

University of Groningen

Structural analysis of the SDSS Cosmic Web

Platen, Erwin; Weygaert, Rien van de; Jones, Bernard J.T.; Vegter, Gert; Aragón Calvo, Miguel A.

Published in:
Monthly Notices of the Royal Astronomical Society

DOI:
[10.1111/j.1365-2966.2011.18905.x](https://doi.org/10.1111/j.1365-2966.2011.18905.x)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2011

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Platen, E., Weygaert, R. V. D., Jones, B. J. T., Vegter, G., & Aragón Calvo, M. A. (2011). Structural analysis of the SDSS Cosmic Web: I. Non-linear density field reconstructions. *Monthly Notices of the Royal Astronomical Society*, 416(4), 2494-2526. <https://doi.org/10.1111/j.1365-2966.2011.18905.x>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Structural analysis of the SDSS Cosmic Web – I. Non-linear density field reconstructions

Erwin Platen,¹ Rien van de Weygaert,^{1*} Bernard J. T. Jones,¹ Gert Vegter²
and Miguel A. Aragón Calvo³

¹Kapteyn Astronomical Institute, University of Groningen, PO Box 800, 9700 AV, Groningen, the Netherlands

²Bernoulli Institute for Mathematics and Computer Science, University of Groningen, PO Box 407, 9700 AK Groningen, the Netherlands

³Department of Physics and Astronomy, The Johns Hopkins University, 3400 North Charles Street, Baltimore, MD 21218, USA

Accepted 2011 April 11. Received 2011 March 19; in original form 2010 June 9

ABSTRACT

This study is the first in a series in which we analyse the structure and topology of the Cosmic Web as traced by the Sloan Digital Sky Survey (SDSS). The main issue addressed in the present study is the translation of the irregularly distributed discrete spatial data in the galaxy redshift survey into a representative density field. The density field will form the basis for a statistical, topological and cosmographic study of the cosmic density field in our Local Universe.

We investigate the ability of three reconstruction techniques to analyse and investigate web-like features and geometries in a discrete distribution of objects. The three methods are the linear Delaunay Tessellation Field Estimator (DTFE), its higher order equivalent Natural Neighbour Field Estimator (NNFE) and a version of the Kriging interpolation adapted to the specific circumstances encountered in galaxy redshift surveys, the *Natural Lognormal Kriging* technique. DTFE and NNFE are based on the local geometry defined by the Voronoi and Delaunay tessellations of the galaxy distribution.

The three reconstruction methods are analysed and compared using mock magnitude- and volume-limited SDSS redshift surveys, obtained on the basis of the Millennium simulation. We investigate error trends, biases and the topological structure of the resulting fields, concentrating on the void population identified by the Watershed Void Finder. Environmental effects are addressed by evaluating the density fields on a range of Gaussian filter scales. Comparison with the void population in the original simulation yields the fraction of false void mergers and false void splits.

In most tests DTFE, NNFE and Kriging have largely similar density and topology error behaviour. Cosmetically, higher order NNFE and Kriging methods produce more visually appealing reconstructions. Quantitatively, however, DTFE performs better, even while being computationally far less demanding. A successful recovery of the void population on small scales appears to be difficult, while the void recovery rate improves significantly on scales $>3 h^{-1}$ Mpc. A study of small-scale voids and the void galaxy population should therefore be restricted to the local Universe, out to at most $100 h^{-1}$ Mpc.

Key words: methods: data analysis – methods: numerical – methods: statistical – cosmology: observations – large-scale structure of Universe.

1 INTRODUCTION

Over the past 30 yr, a clear paradigm has emerged as large redshift surveys opened the window on to the distribution of matter in our Local Universe: galaxies, intergalactic gas and dark matter exist in a wispy web-like spatial arrangement consisting of dense compact

clusters, elongated filaments, and sheet-like walls, amidst large near-empty void regions, with similar patterns existing at higher redshift, albeit over smaller scales. The Cosmic Web is the fundamental spatial organization of matter on scales of a few up to a hundred megaparsec, scales at which the Universe still resides in a state of moderate dynamical evolution (Zel'Dovich 1970; Peebles 1980; Bond, Kofman & Pogosyan 1996). Its appearance has been most dramatically illustrated by the recently produced maps of the nearby cosmos, the 2dF galaxy redshift survey (2dFGR), the Sloan Digital

*E-mail: weygaert@astro.rug.nl

Sky Survey (SDSS) and the two-micron all-sky survey redshift surveys (Colless et al. 2003; Tegmark et al. 2004; Huchra et al. 2005).

According to the standard lore of structure formation, structures emerged from small perturbations in the primordial field of Gaussian density and velocity perturbations. Under the force of gravity these fluctuations grow and cluster to become the present-day observed structures. At large scales the density field has been evolving (quasi)-linearly and still retains much information on cosmological parameters and structure formation. The linear density field provides an abundance of probes from which cosmological information can be extracted. Prominent probes are the clustering of galaxies, which has been used to infer the underlying primordial power spectrum of density fluctuations (Peebles 1980; Percival et al. 2001; Tegmark et al. 2004), the temperature and polarization anisotropies in the cosmic microwave background (e.g. Smoot et al. 1992; Spergel et al. 2003) and the shearing of galaxy images by the gravitationally lensed photon paths through the inhomogeneous matter distribution (Mellier 1999; Massey et al. 2007; Hoekstra & Jain 2008).

While these (quasi)-linear cosmological probes have yielded an impressive amount of cosmological information, the exploitation of the pronounced non-linear patterns of the Cosmic Web towards probing cosmological parameters and cosmic structure formation has been less fortuitous. Even though the morphology, shape and other statistical characteristics of the quasi-linear cosmological density field forms a direct reflection of the structure assembly process in the Universe, on the corresponding small-scales non-linear growth has significantly altered and erased some of the essential cosmological information. The absence of an objective and quantitative procedure for identifying and isolating clusters, filaments and voids in the cosmic matter distribution has been a major obstacle in investigating the structure and dynamics of the Cosmic Web. The overwhelming complexity of the individual structures and their connectivity, the huge range of densities and the intrinsic multi-scale nature prevent the use of simple tools that may be sufficient in less demanding problems. However, various interesting new approaches and methods have been forwarded in the past few years, often based on ideas stemming from image processing, mathematical morphology and medical imaging (e.g. Aragón-Calvo et al. 2007, 2010; Genovese et al. 2010; Sousbie 2011; Way, Gaziz & Scargle 2011).

This study is the first in a series in which we systematically investigate the web-related structures found in the SDSS. It involves a systematic program in which we explore the cosmography of the Local Universe, assess the statistical characteristics of the density field, identify and categorize the voids and filaments within the SDSS galaxy redshift sample, and study the biasing of the galaxy population with respect to the mass distribution and the dependence of galaxy properties on the large-scale environment.

The intention of this paper is the reconstruction of the underlying (non-linear) density field by translating the spatially irregularly distributed and discrete galaxy sample in the SDSS GRS into a representative density field. The density field reconstructions described in this study will form the basis of a series of studies in which we address the statistical and topological properties of the cosmic density field in our Local Universe and in which we analyse the cosmography, void population and spinal structure of the local Cosmic Web in the SDSS. The cosmographic description of the nearby Universe includes the identification and cataloguing of the filaments, voids and clusters in the SDSS galaxy distribution. The key motivation for the density field reconstruction techniques should therefore be that it allows us to probe the complex quasi-linear (and non-linear) structures that we find in GRS. This involves the ability to repro-

duce the distinct anisotropic – filamentary and wall-like – features that make up the Cosmic Web, as well as the ability to trace the hierarchical substructure of the cosmic matter distribution and the ability to identify the void population that forms one of its most salient features.

We are specifically interested in the performance of the fast and efficient linear tessellation-based Delaunay Tessellation Field Estimator (DTFE) density field reconstruction technique (Schaap & van de Weygaert 2000; Schaap 2007; van de Weygaert & Schaap 2008). In addition, we investigate two additional higher order techniques which are potentially suited for representing the quasi-linear density field of the Cosmic Web, the local natural neighbour interpolation [Natural Neighbour Field Estimator (NNFE)] and a non-local Kriging interpolation technique, the Natural Lognormal Kriging formalism. By means of an extensive comparison we evaluate which aspects of the density field are best reproduced by either DTFE, NNFE or Natural Lognormal Kriging, and which of these methods is best suited to function for further structural analysis.

In order to be able to assess the reliability of the results of our structural tools, it is crucial to understand the details and errors in reconstructed density fields. We will therefore present a detailed comparison study between three different reconstruction methods, and assess their density and topological errors over a range of scales. This will range from the small non-linear scales, at $1 h^{-1}$ Mpc, to a scale of $10 h^{-1}$ Mpc, which represents the transition from quasi-linear to the linear regime.

We focus on the translation of the galaxy positions in the SDSS DR6 GRS to a representative density field within the survey volume. Currently, SDSS encloses the largest and deepest contiguous region of the nearby Universe mapped by a GRS. This makes the SDSS an ideal data sample for a full three-dimensional density field reconstruction. A first impression of the resulting DTFE density field map of a region in the SDSS DR6 survey is shown in Fig. 1. Our techniques will be generally applicable to any uniform GRS. Even though survey limits and scales are different from that of the SDSS, it will be straightforward to carry over the results to other redshift surveys.

1.1 Inferring cosmic density fields

The richest source of data for investigating the intricate web-like cosmic matter distribution is the distribution of galaxies in GRSs. We shall make the assumption that the galaxy distribution is a representative tracer of the underlying mass density field and of the underlying structure. Different samples selected from such catalogues may of course trace different structures.

We set out to explore high fidelity methods for reconstruction of the density field from the discrete galaxy distribution. We require accurate local density estimates that can be used to reliably compute structural and topological indicators. Computational efficiency is an important factor since our reconstruction technique has to be able to deal with galaxy samples of a million of galaxies as well as with N -body simulations comprising orders of magnitude more particles.

There are two immediate and important implications and complications with respect to our intention of including the (quasi)-non-linear components in the density field. The first one is that we have to be aware of the noise components that tend to be included in the reconstructions as well (Schaap 2007; van de Weygaert & Schaap 2008). The second one is that non-linear data are typically not well behaved, marked by strong gradients in the density field. It means that we have to take special care of those locations marked by non-linearities in the data.

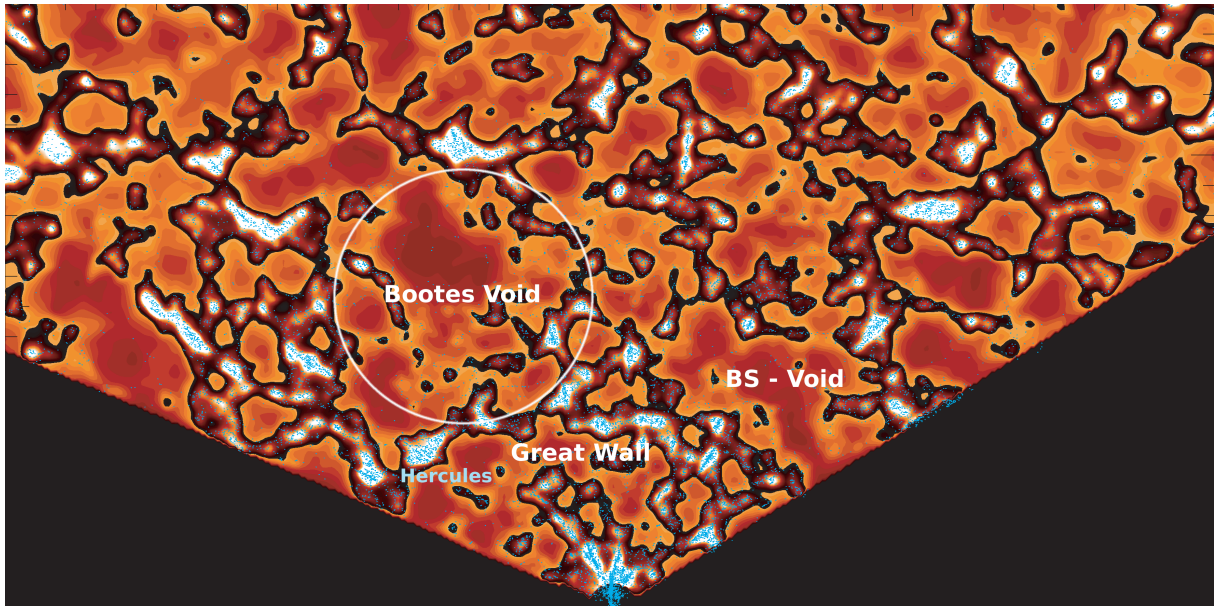


Figure 1. A visualization of the (DTFE) SDSS density field ($1 h^{-1}$ Mpc); the contour levels are divided roughly into the overdense and the underdense regime. Both the galaxies (blue dots) and the density represent a slice of $12 h^{-1}$ Mpc thickness. Some of the most prominent features have been named, like the *Bootes SuperVoid* (Kirshner et al. 1981) and the large supervoid (*BS SuperVoid*) identified by Bahcall & Soneira (1982). Also the largest overdense structure, the (*Coma*) *Great Wall*, and the location of the *Hercules* supercluster within the Wall are indicated.

1.2 Spatial point processes and continuous density fields

We pursue the reconstruction of a density field from a spatial point process consisting of irregularly distributed points. We assume that the local intensity of the points is a fair tracer of the density and that these values are samples of a continuous underlying density field. In doing so we appreciate that differently defined samples may trace different structures. The reconstruction problem we address here is a data processing problem. The interpretation of the results is an astronomical process.

When the spatial point process is defined by the galaxy distribution, the situation will be more complex. The Cosmic Web is marked by a distinct luminosity, colour and morphology segregation. A strong and systematic trend of galaxy morphology with density has been established a few decades ago by Dressler (1980). Early-type galaxies preferentially in rich groups and clusters and late-type galaxies residing mainly in filaments and walls (e.g. Giovanelli, Haynes & Chincarini 1986). Other strong systematic clustering trends with luminosity and colour have also been established, such as in the complete SDSS DR7 galaxy sample (Zehavi et al. 2010). A fair sample of galaxies should in principle take this intrinsic segregation into account. We are pursuing this in a forthcoming publication. In this study, mainly intent on establishing our reconstruction technology, we will consider the total galaxy population. The reconstructed density maps are therefore unlikely to represent a fair reflection of the underlying dark matter network, but will none the less convey the overall pattern of the large-scale structure.

The reconstruction consists of two fundamental steps. The first is the estimation of the local galaxy density. The second is the interpolation of these density values to obtain a continuous spatial density field.

Following straightforward grid-based interpolation methods and more sophisticated adaptive filter techniques, there has been a substantial investment into developing more advanced techniques. Ex-

amples of recently introduced techniques to follow the multi-scale nature of the Cosmic Web is the Multi-scale Morphology Filter (Aragón-Calvo et al. 2007) and the Hierarchical SpineWeb technique (Aragón-Calvo et al. 2010), a Morse theory-based formalism that is closely related to the Watershed Void Finder (WVF) by Platen et al. (2007). An example of an alternative route involves the wavelet reconstruction by Martínez et al. (2005).

Here, we concentrate in particular on the density estimation and the subsequent field interpolation, in anticipation of the post-processing and feature detection studies of the SDSS DR6/DR7 survey, the subject of the additional papers in this series. To deal with the shot noise and necessary selection effects in the resulting density fields, the analysis will involve filtering, a necessary post-processing step.

1.2.1 Density estimation

The first step in the reconstruction procedure is density estimation. We show how this fits into our general scheme in Fig. 2. A similar, but slightly different, outline can be found in Kitaura & Enßlin (2008).

In the literature we may find a large variety of density estimators (e.g. Silverman 1986; Katkovnik 2002; Sain 2002; Miller 2003). Usually, they involve filters of some kind. Dependent on the filter kernel, density estimates may have a local character or include the information from a wider range of points in a point distribution. Filter kernels may have a rigid scale and shape, or they may be adapting themselves to the local point density. An example of a global estimator is a Gaussian kernel, which takes along the information – be it weighted – from distant points. A well-known example of rigid local estimators are the CIC and TSC grid interpolation formalisms (e.g. Hockney & Eastwood 1981). A considerably more flexible and adaptive filter kernel, frequently used in current N -body studies, is that of the spline-based interpolation recipes used in Smooth

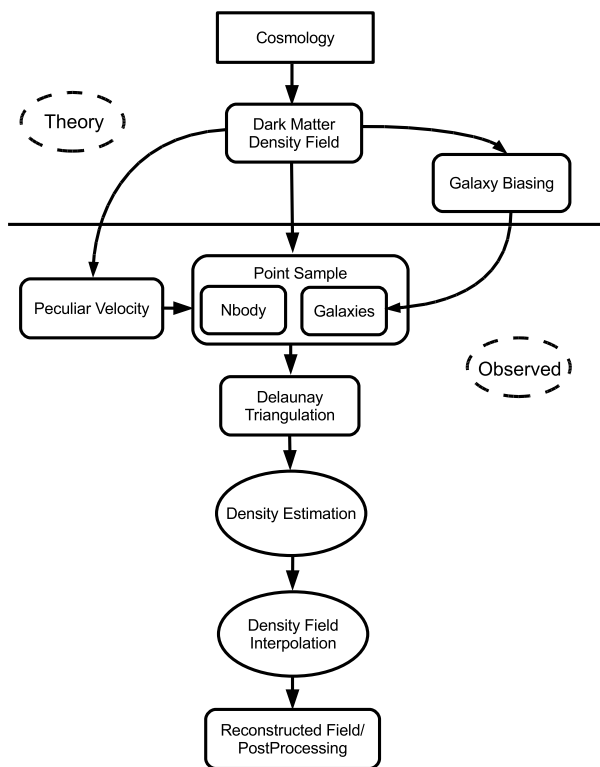


Figure 2. The schematic outline of our cosmological reconstruction procedure. The *Point Sample* is discussed in Section 2. The *Density Estimation* item will be treated in Sections 1.2.1 and 3 and the *Density Interpolation* methods are introduced in Sections 4.1, 4.2 and 4.3. Sections 6, 7 and 8 deal with the *Post-Processing*.

Particle Hydrodynamics codes (e.g. Monaghan 1992) whose scale is determined by the distance to the K th nearest neighbour.

Recently, various local and adaptive density estimators have been shown to provide highly favourable results for complex cosmological matter distributions, marked by a large dynamics range of scales and density values and intricate geometric patterns. One of these exploits the adaptivity of the kD tree (Ascibar & Binney 2005; Sharma & Steinmetz 2006). We also note the use of the Epanechnikov kernel estimator in the identification of superclusters (Einasto et al. 2007).

For the three reconstruction techniques addressed in this study, we use the local DTFE density estimate introduced by Schaap & van de Weygaert (2000) (see Appendix B). It sets the density value at a sample point proportional to the inverse volume of the contiguous Voronoi cell around a sample point, i.e. the sum of the Delaunay tetrahedra to which the sample point belongs (Schaap & van de Weygaert 2000; Schaap 2007; van de Weygaert & Schaap 2008). Tessellation-based methods are based on the realization that the optimal estimate for the spatial density at the location of a point \mathbf{x}_i in a discrete point sample \mathcal{P} is given by the inverse of the volume of the corresponding Voronoi cell (Okabe et al. 2000). They have been introduced by Brown (1965) and Ord (1978) and were first used in astronomy, for the specific purpose of devising source detection algorithms, by Ebeling & Wiedenmann (1993).

In an extensive study, Schaap (2007) demonstrated that the DTFE estimates are substantially better than those of the rigid grid-based CIC and TSC techniques. Also, it outperforms the adaptive smooth particle hydrodynamics (SPH) spline performance, in particular in areas with substantial density gradients.

1.2.2 Interpolation

Interpolation on randomly scattered points aims to approximate a continuous function constrained by the available data points. A wide variety of approaches have been put forward, each with their own advantages and disadvantages. These methods can be roughly divided into two categories, global and local methods. Local methods have the benefit that they are fast and able to deal with large data sets. Global methods tend to produce smoother interpolated functions but are computationally more expensive.

An overview and discussion of spatial interpolation techniques can be found in Press (2007) (chapter 3.4) and in Watson (1992). We refer to Lombardi (2002) for a detailed review of the statistical properties of a large number of techniques, and to Franke (1982) and Amidror (2002) for a detailed comparison between various methods.

Amongst the more sophisticated interpolation techniques we can distinguish various classes: inverse distance-based methods (IDW), moving least-squares methods, the class of radial basis function (RBF) interpolation techniques, Kriging interpolation methods (Kriging 1951; Matheron 1963, see Section 4.3) and triangulation-based methods, both the linear DTFE technique (Schaap & van de Weygaert 2000; van de Weygaert & Schaap 2008, Section 4.1) and the higher order NNFE natural neighbour interpolation technique (Sibson 1981; Watson 1992, Section 4.2).

RBF and Kriging interpolation methods use predefined kernels to interpolate the field. The kernel in RBF methods is a basis function that spans the space of all the interpolating functions. Kriging interpolation uses spatial correlations between sample points, with its kernel being equal to the corresponding covariance function. The method, introduced by Matheron (1963), is the *best linear unbiased estimator* of a density value given a set of measured sample points at irregularly spaced points. Given the commonly accepted fact that the primordial density perturbation field is Gaussian, we may therefore see the Kriging interpolator as the natural choice for reconstructing the cosmological density fields which emerged out of these primordial Gaussian circumstances.

There is a distant relationship of the Kriging interpolation to Wiener filtering techniques (Wiener 1949; Rybicki & Press 1992). However, Wiener filtering is based on a different philosophy than Kriging, in that it includes a model for the noise and is evaluated in Fourier space. Also, classical Wiener filtering is predicated on an underlying Gaussian distribution. As a result, it has the serious disadvantage of suppressing or substantially diluting non-linear structures of interest. More advanced recent developments and applications of Wiener filters to the reconstruction of the density distribution have largely remedied its capacity for reconstructing the density distribution (Kitaura & Enßlin 2008; Kitaura et al. 2009; Kitaura, Jasche & Metcalf 2010).

A rather different approach is advocated in the triangulation-based interpolation techniques. Both the linear DTFE (Schaap & van de Weygaert 2000; Schaap 2007; van de Weygaert & Schaap 2008) and the higher order NNFE (Sibson 1981; Watson 1992; Braun & Sambridge 1995) use the neighbourhood relationships defined by the Voronoi and Delaunay tessellation of the point sample to establish a fully adaptive, irregular and local interpolation grid. DTFE uses the Delaunay triangulation to reconstruct in a self-adaptive, mass conservative and parameter free way the underlying spatial (density) distribution. In combination with the DTFE tessellation-based sample points density estimates (see previous Section 1.2.1), the DTFE interpolation leads to a volume-covering density field which has been shown to recover the hierarchical as well as the

anisotropic morphology of the Cosmic Web (Schaap 2007; van de Weygaert & Schaap 2008).

1.3 DTFE, NNFE and Kriging

The DTFE method will be compared to two other techniques having a similar potential for a proper reconstruction of the Cosmic Web. The NNFE *Natural Neighbour Interpolation* technique shares the local nature of DTFE, but involves higher order interpolations. The third formalism is *Natural Lognormal Kriging*, which is a version of the Kriging interpolation, and thus involves a non-local higher order interpolation methodology.

The crucial step of importance in our investigation is the interpolation step (see diagram Fig. 2). Using the same initial sample point density estimates, the first step of the reconstruction procedure, we can assess the relative merits of DTFE, NNFE and Kriging interpolators.

1.4 Outline of this study

We start by describing the data samples in Section 2, the SDSS survey sample and a mock galaxy sample which mimics the SDSS, obtained from the Millennium simulation. The local DTFE density estimate at the location of each of the sample galaxies is the subject of Section 1.2.1. In the subsequent section we present and describe each of the three interpolation techniques investigated in this study, DTFE in Section 4.1, NNFE in Section 4.2 and Lognormal Kriging in Section 4.3. The comparison throughout the paper is based on one specific sample, the density field reconstruction of the SDSS mock galaxy sample. Following a discussion in Section 5 of the qualitative appearance of the density maps by each of the three methods, we turn to an intensive quantitative error and quality analysis of the density field in Section 6. In Section 7, this is followed by an investigation of topological structure of the web-like galaxy distribution in the survey, mainly based on the void population as traced by the WVF. Section 8 presents the density field reconstruction of the SDSS data sample, and in Section 9 we summarize this study.

Following this first paper in a series will be a statistical study of the density field, focusing in particular on the one-point probability density function. Subsequently, we will present and discuss the cosmography of the reconstructed local Universe. Later studies will analyse the void population in the SDSS density field, and concentrate on the properties of galaxies as a function of the large-scale environment as characterized by our technique.

2 THE DATA

Our study is based on two major data sets. The principal one is the genuine SDSS DR6 GRS. For the purpose of understanding the errors and artefacts in the density field reconstruction we use a set of galaxy mock catalogues that model this SDSS data set. The mock samples are obtained from the Millennium simulation (Springel et al. 2005).

2.1 The SDSS galaxy sample

For our analysis we use the main galaxy sample in the North Galactic Cap from the sixth data release of the SDSS (York et al. 2000; Strauss et al. 2002; Stoughton et al. 2002; Adelman-McCarthy et al. 2008).

The SDSS DR6 consists of various contiguous regions. We restrict ourselves to the largest contiguous region, the northern strip of the North Galactic Cap (NGC) region. The sample was retrieved from the SDSS ‘casjobs server’ (www.sdss.org) using the SQL query interface. Relevant properties were downloaded and most of the post-processing was done on a workstation. We did not attempt to assign galaxies with missing redshift due to fibre collisions. This gives a lower density estimated at the position of the missing redshift. However, given the sampling scheme used to resample the density field to a regular grid, the probability of sampling one of the affected areas is very low. The problem occurs only in the high-density regions of the catalogue.

The spectroscopic SDSS galaxy sample is almost complete between a Petrosian magnitude limit of $m_r = 14.5$ and $m_r = 17.77$. We assume that the completeness of the sample does not vary significantly, even though there are in fact some angular variations (Blanton et al. 2003).

For our study, we select the SDSS galaxies that are located within a comoving box of $600 h^{-1}$ Mpc. In terms of the survey coordinates (X,Y,Z) (see Appendix A for definition), the observer is located at (X,Y,Z) = (300, 0.300) h^{-1} Mpc and the centre of the northern strip is rotated to lie parallel to the Y-axis starting at (X, Z) = (300, 300) h^{-1} Mpc. In Fig. 3 the galaxies are plotted in the XY-projection (top) and the XZ-projection (bottom).

Within the (600 h^{-1} Mpc)³ volume, there are a total of 311 474 galaxies with a magnitude less than the magnitude limit $m_r = 17.77$.

2.2 The SDSS mock samples

The Millennium simulation (Springel et al. 2005) was used to construct galaxy mock catalogues which emulate the SDSS galaxy redshift sample. They are needed to get estimates of the errors induced by the magnitude selection, redshift distortions as a result of peculiar velocities and errors resulting from the survey mask.

We use the semi-analytical galaxy samples of De Lucia & Blaizot (2007), Bower et al. (2006) and Bertone, De Lucia & Thomas (2007) to construct our own mock samples. They are less detailed as the mock catalogue generated by Blaizot et al. (2005); they are perfectly suited for a representation of all necessary aspects of the SDSS sample.

The same (X,Y,Z) coordinate system and box size of 600 h^{-1} Mpc have been used for the mock galaxy samples. The mock galaxy catalogues are constructed as follows.

- (i) Periodically tile the Millennium cube to obtain enough cosmic volume.
- (ii) Calculate the redshift of the model galaxies with respect to the observer.
- (iii) Compute the apparent magnitude of each model galaxy from its absolute magnitude and redshift.
- (iv) Select the model galaxies brighter than $m_r = 17.77$.
- (v) Add the peculiar velocity to the Hubble redshift to obtain the total redshift.
- (vi) Apply the observational mask of the DR6-NGC sample to decide whether the model galaxy is included in the mock catalogue.

2.3 Redshift-space distortions

In this study we assess redshift-space density maps as well as partially corrected ‘real’ space density maps.

Redshift-space surveys like the SDSS are beset by distortions in the estimated distance. The result of large coherent cosmic flows,

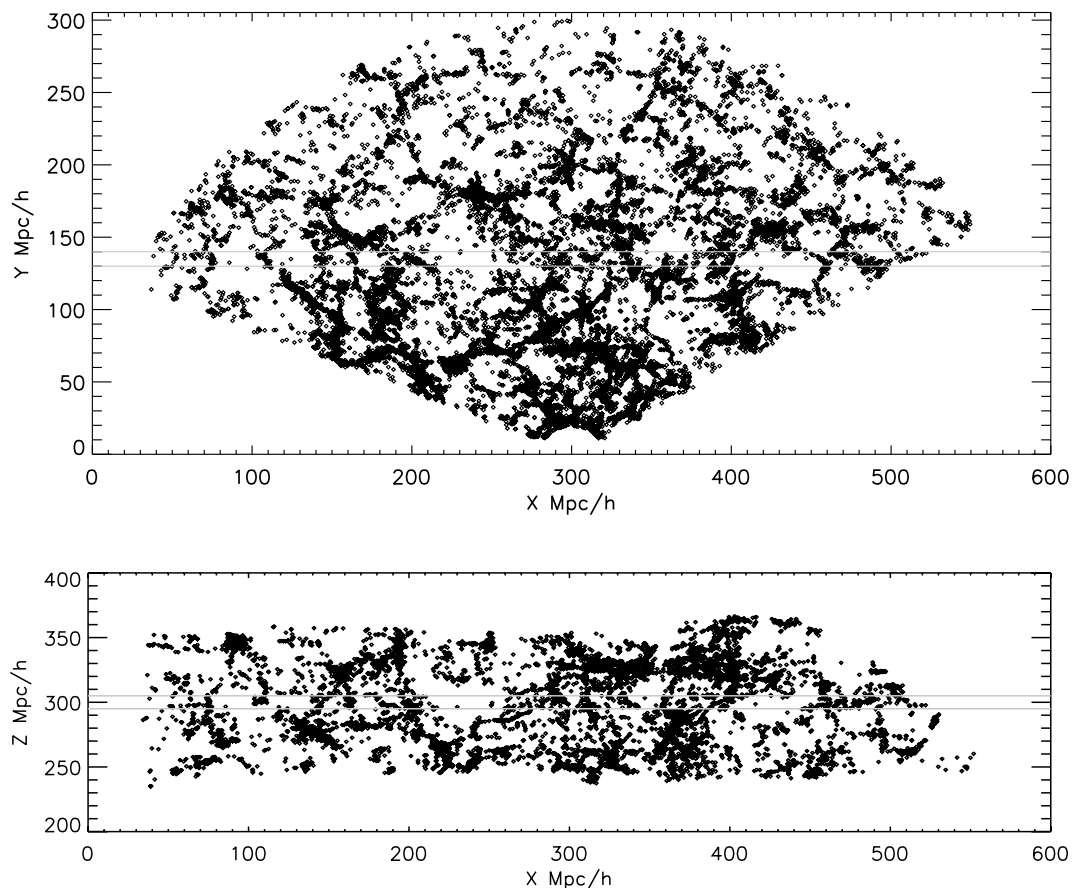


Figure 3. The top figure shows the SDSS galaxies in the X and Y coordinates out to a distance of $300 h^{-1}$ Mpc. The lower figure shows an XZ slice perpendicular to the XY plane at $Y = 135 h^{-1}$ Mpc. The corresponding boundaries of both slices are indicated in the other figure, e.g. grey lines in the top figure show the Y limits of the bottom figure and vice versa. Both slices have a thickness $10 h^{-1}$ Mpc.

infall velocities on to clusters and highly non-linear ‘thermal’ velocities within clusters, these redshift distortions can have a dramatic effect on the estimated distances and reshape the large-scale matter distribution. However, for our purposes here, which is testing and comparing the methods, the presence or absence of *fingers of God* is immaterial.

Hence, in this paper, which deals with the mock SDSS catalogues, we do not correct for the redshift distortions induced by large-scale coherent cosmic flows. In particular the DTFE density reconstruction is locally adaptive and so the reconstruction is not corrupted by the presence of elongated radial features; this can be seen in Fig. 4.

The redshift distortions will, however, be addressed in a following paper in which we analyse the real SDSS catalogue.

2.4 Magnitude- versus Volume-limited Samples

For the analysis of the SDSS sample, we extract two different samples from the full SDSS galaxy sample. These are a volume-limited (i.e. absolute magnitude limited) sample and an apparent magnitude-limited sample. Each sample is used for different aspects of our analysis.

Volume-limited sample

A *volume-limited* galaxy sample is defined in order to assure a uniform galaxy coverage over the survey volume. A volume-limited

sample consists of a subset of galaxies which are homogeneously sampled throughout the sample volume. It has the advantage that each sample galaxy is an equal weight tracer of the underlying density field, and the resulting field will be statistically uniform. Our volume-limited sample has a distance limit of $300 h^{-1}$ Mpc and includes all galaxies brighter than $M_r < -20.45$, roughly representing the galaxies brighter than L_* .

While the uniformity of the volume-limited samples assures a straightforward error assessment for any analysis, it has the disadvantage of losing the high spatial resolution represented by fainter galaxies nearby, therefore it does not necessarily have the *smallest* error.

Magnitude-limited sample

The *magnitude-limited* sample contains all 311 474 SDSS galaxies brighter than $m_r = 17.77$. While a magnitude-limited sample takes along all sampled information, one needs to correct for the inhomogeneous selection process.

A characteristic of magnitude-limited surveys is the change of intrinsic spatial resolution as we proceed out to larger distances. Potentially, this could be a serious issue when galaxies would be biased in a very complicated fashion (e.g. higher order biasing). This would render it very difficult to infer the density field at large distances, where only the most luminous objects would remain visible. By default, we therefore assume that all galaxies – independent of their luminosity – are a fair tracer of the density field.

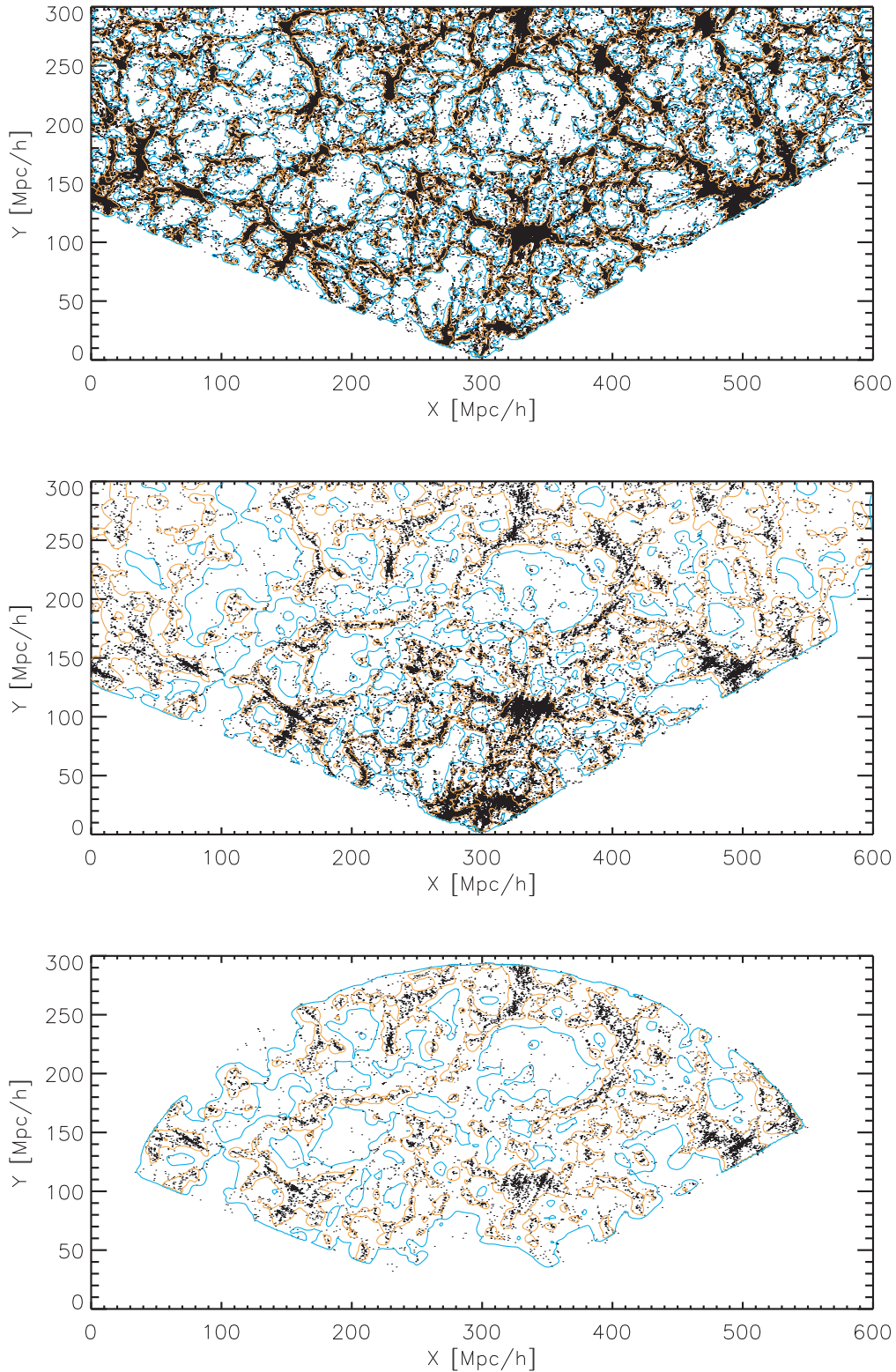


Figure 4. SDSS mock catalogues from the Millennium simulation. Top: full Millennium catalogue. Centre: magnitude-limited mock catalogue. Bottom: volume-limited mock catalogue. The mock galaxies are indicated by black dots. The blue and orange contour lines are NNFE density contour lines at density contrast $\rho/\rho_u = 0.25$ (blue) and 1.0 (orange). For reasons of clarity, the top slice has a thickness of $6.0 \ h^{-1} \text{ Mpc}$ while the magnitude and volume-limited slices each $12.0 \ h^{-1} \text{ Mpc}$: showing a thinner slice for the two last samples would show too few galaxies.

Following this assumption, we correct for the dilution as a function of survey depth by weighing each sample galaxy by the reciprocal $w(z)$ of the radial selection function $\psi(z)$ at the distance of the galaxy. For the SDSS, the selection function $\psi(z)$ as a function of redshift z is well fitted by the expression forwarded by Efstathiou & Moody (2001)

$$\psi(z) = \exp \left\{ - \left(\frac{z}{z_r} \right)^\beta \right\}, \quad (1)$$

where z_r is the characteristic redshift of the distribution and β specifies the steepness of the curve. The corresponding number density $N(z)$ of galaxies at redshift z is

$$N(z)dz = Az^2\psi(z)dz \\ = Az^2 \exp \left\{ - \left(\frac{z}{z_r} \right)^\beta \right\} dz, \quad (2)$$

where A is a normalization constant and the z^2 term represents the increase of volume as a function of z . The resulting galaxy weights $w(z)$ are

$$w(z) = 1/\psi(z) = \exp \left\{ \left(\frac{z}{z_r} \right)^\beta \right\}. \quad (3)$$

When incorporating $w(z)$, the weights into the density field reconstruction, the normalization of the resulting density field may be achieved by modelling the details of the selection function and calculating the appropriate normalization constant. We chose to follow the alternative of calculating the average of the reconstructed density field and subtracting it from the reconstruction. This is a simple and straightforward procedure, and perfectly valid as long as the volume of the sample is large enough to enable the estimate of the true average.

Mock samples

From the mock catalogues we select three different samples. The *full mock catalogue* comprises all Millennium (semi-analytical) model galaxies within the DR6-NGC mask. The *magnitude-limited* mock sample includes all model galaxies resulting from the procedure described above. The third sample is a *volume-limited* set with an absolute magnitude $M_r < -20.45$.

To appreciate the differences between the *magnitude-limited* and the *volume-limited* galaxy sample, and also the *full mock catalogue* within the SDSS volume, the three mock catalogues are shown in Fig. 4.

The contours superimposed on the corresponding galaxy distributions are the resulting NNFE density field contours (see Section 4.2).

3 LOCAL DTFE DENSITY ESTIMATE

Throughout this study, we use the local DTFE density estimate, following the definition by Schaap & van de Weygaert (2000). In Appendix B one may find more details of the DTFE procedure which we followed (for an extensive review see van de Weygaert & Schaap 2008). Implicitly, and for simplicity, we assign to each sample galaxy the same mass m_i , i.e. the density value is predicated on the number density of galaxies.

The sample point DTFE density value is inversely proportional to the volume of the local neighbourhood as defined by the Voronoi tessellation of the spatial galaxy distribution. Schaap & van de Weygaert (2000) argued that the inverse of the volume of the *contiguous Voronoi cell* is the proper density estimate, assuring mass

conservation for the subsequent linear interpolation step. The contiguous Voronoi cell, sometimes dubbed *umbrella* in the computational geometry literature, is the region defined by all Delaunay tetrahedra of which a given sample point is a vertex and which it shares with its *natural neighbours*. A two-dimensional illustration of a contiguous Voronoi cell of a point is shown as the surrounding hatched region in the left-hand frame of Fig. 5.

The density value at each sample point is determined following the construction of the Delaunay triangulation (see e.g. Delaunay 1934; Aurenhammer 1991).¹ Within the triangulation we identify for each sample point i all N_i neighbouring tetrahedra \mathcal{T}_j (Fig. 5), which together constitute the contiguous Voronoi cell $\mathcal{W}_i \cup_j \mathcal{T}_j$. Summation of the individual tetrahedral volumes $V(\mathcal{T}_j)$ yields the volume of the contiguous Voronoi cell,

$$V(\mathcal{W}_i) = \sum_{j=1}^{N_i} V(\mathcal{T}_j). \quad (4)$$

For the three-dimensional SDSS sample volume, the resulting DTFE estimate of the density \hat{f}_i at sample point i is (see equation B2)

$$\hat{f}_i = \frac{4 w(z_i)}{V(\mathcal{W}_i)}, \quad (5)$$

where the weight $w(z_i)$ is the sample selection weight at the galaxies' redshift z_i (equation 3). Note that the factor four takes account of the fact that in three dimensions each sample point belongs to four tetrahedra. In practice, the density of all particles is calculated by looping in sequence over all Delaunay tetrahedra.

3.1 Shot noise errors

A local density estimator such as the contiguous Voronoi cell has the advantage of being very sensitive to the signal. However, this also implies them to have a high sensitivity to shot noise present in the data.

For appreciating the influence of shot noise in the DTFE density estimates, we turn to the probability distribution of the estimated intensity $\hat{\lambda}$ for a 3D Poisson point process with intensity $\lambda = 1$. Schaap (2007) found that it can be very well approximated by

$$p(\hat{\lambda}) = \frac{1944}{5} \hat{\lambda}^{-8} e^{(-6/\hat{\lambda})}. \quad (6)$$

The first observation is that the DTFE density estimator is unbiased, as the mean of $p(\hat{\lambda})$ is equal to 1. An estimate of the error involved is that of the variance, $\sigma_{\hat{\lambda}}^2 = 1/5$, and is equal to ~ 57 per cent. Evidently, the distribution is non-Gaussian with a long tail extending towards high-density values.

3.2 Centroidal Voronoi tessellations

The high-density tail of the DTFE density estimate, and the implied shot noise level, can be suppressed or regularized by using the centroidal Voronoi tessellation (CVT) (see Lloyd 1982; Browne 2007). For a CVT the generating point distribution is such that the generating points are the mass centres of the resulting Voronoi cells.

The calculation of a CVT is usually done by means of an iterative procedure known as the Lloyd iteration. Starting with an originally random point distribution, the centre of mass of the corresponding Voronoi cells is computed. Subsequently, the points are displaced to these centres. After a sequence of iteration steps, the resulting

¹ The Delaunay triangulation in this work has been computed using the CGAL library, Computational Geometry Algorithms Library (CGAL).

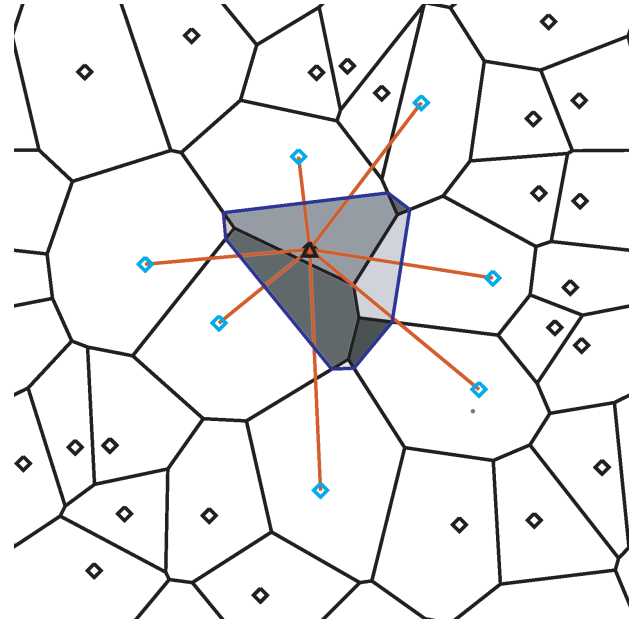
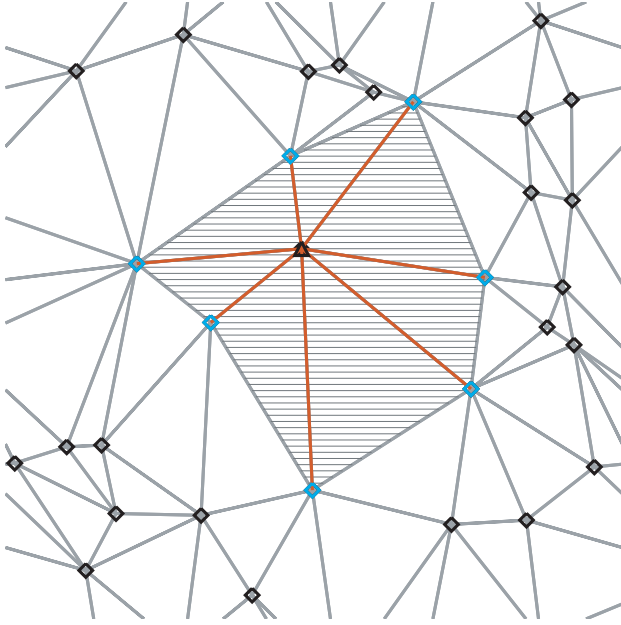


Figure 5. Voronoi tessellations, Delaunay tessellations and Natural Neighbours. Left: the Delaunay triangulation (grey lines) of a point set (circles). For the central point the natural neighbours are indicated by the blue points, along with the corresponding Delaunay edges (red). The hatched region is the contiguous Voronoi cell of the central point, the composite of all Delaunay triangles shared with its natural neighbours. Right: the Voronoi tessellation of a point sample (black diamonds). Relevant for Natural Neighbour interpolation (NNFE): following the insertion of a central point (black triangle), a new Voronoi cell (region enclosed by the dark blue polygonal boundary) is computed, the second-order Voronoi cell. The grey shaded areas are the overlapping regions with the original Voronoi tessellation.

point distribution tends to converge to a proper CVT constellation. Effectively, the points have been repelling each other.

An impression of the CVT iteration procedure can be obtained from Fig. 6. Involving an initial point distribution, the resulting intensity distribution $p(\hat{\lambda})$ for the tessellations obtained af-

ter zero, two, four, six, eight and 10 Lloyd iterations is shown in the right-hand frame. Clearly, a CVT involves a much more regular distribution: after four iterations $p(\hat{\lambda})$ has turned into a narrow and near symmetric distribution whose high-end tail is almost absent (Fig. 7). Potentially, a CVT might therefore

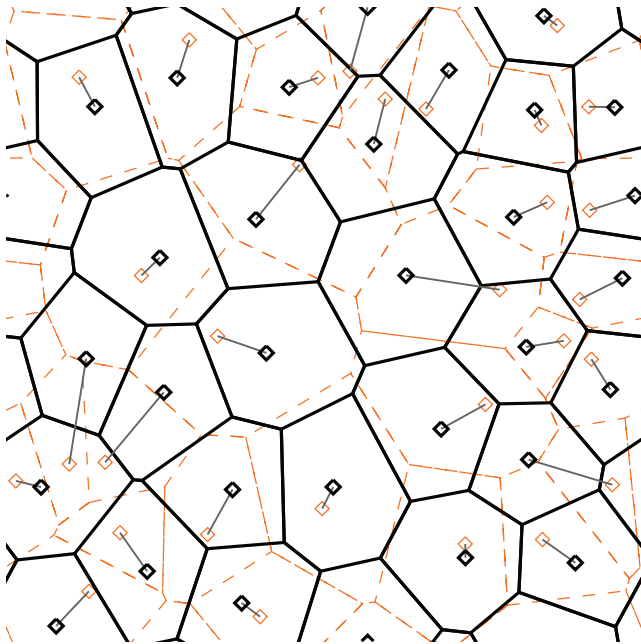


Figure 6. Centroidal Voronoi Tessellation. The (original) Voronoi tessellation for the points of Fig. 5 are plotted in orange. In black we show the CVT with corresponding points after two Lloyd iterations. The displacements are indicated by the grey lines.

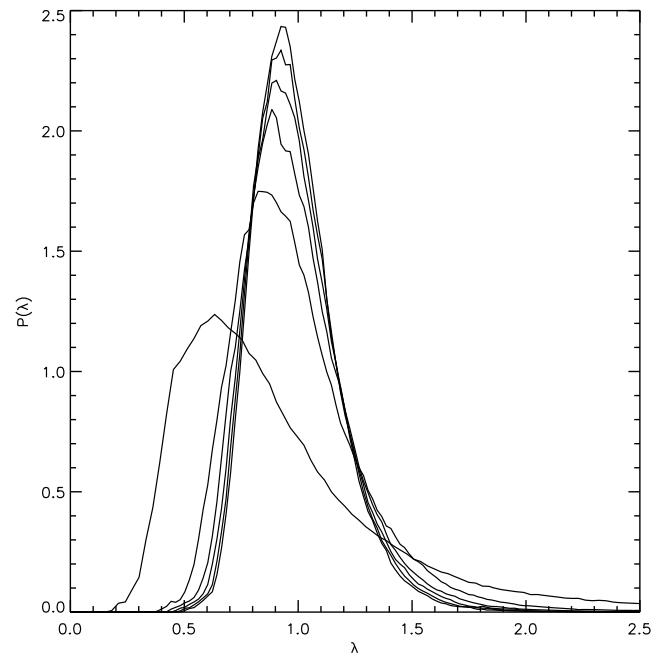


Figure 7. The error probability distribution of the CVT-based galaxy density estimator for a homogeneous Poisson sample. Arranging the curves from broadest to the most narrow they respectively correspond to zero, two, four, six and 10 Lloyd iterations.

help to suppress the density estimate error and its asymmetric distribution.

3.3 DTFE noise: a case study

A visual impression of the shot noise involved in the density estimate is provided in Fig. 8. It concerns a random sample of 4500 points distributed according to an anisotropic Gaussian distribution. This configuration is more representative of what may be expected in the real galaxy distribution.

Densities were estimated according to the pure DTFE procedure (see equation 6) and following a Lloyd CVT procedure of five iterations. The top row of Fig. 8 shows the density contours of the original peak and the raw DTFE and DTFE/CVT field reconstructions. The raw DTFE reconstruction (centre) is highly irregular compared to the original contours (left), while the CVT contours are much more regular (right). The linear profile along the y -axis (at $x = 0.5$, bottom-left panel) emphasizes the visual impression: the DTFE profile (orange) is marked by salient peaks which reflect the high-density tail of $p(\hat{\lambda})$, while the DTFE/CVT

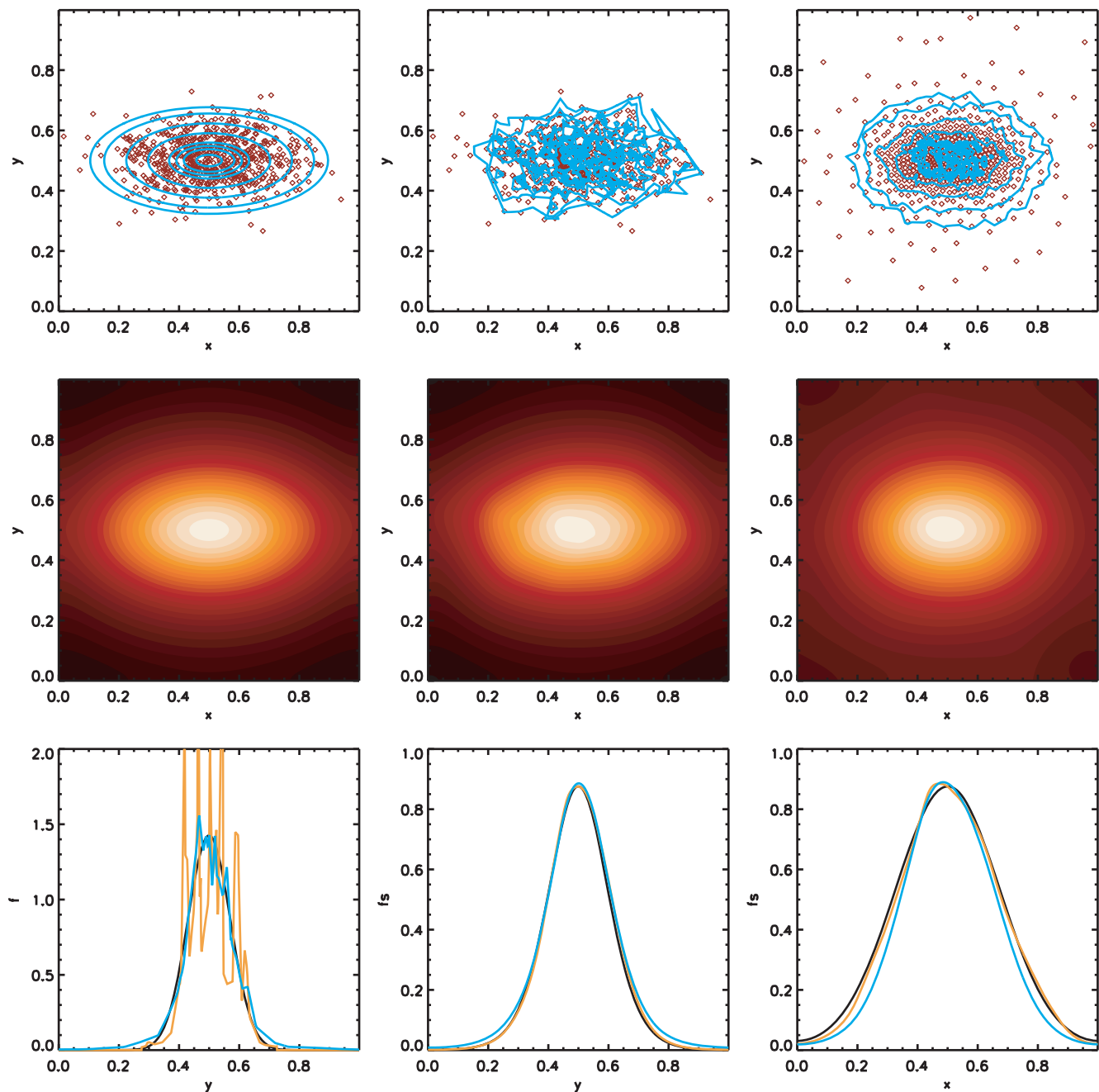


Figure 8. DTFE and DTFE/CVT reconstruction of an anisotropic Gaussian peak. Sampling the peak by 4500 random points, the top row shows the contour levels of the reconstructions: original (left), DTFE (centre) and DTFE/CVT with five Lloyd iterations. The resulting density maps after Gaussian smoothing, with $R_f = 0.05$, is shown in the central row. The bottom row shows linear density profiles through the resulting mass distribution. Left: linear profiles along the y -axis at $x = 0.5$, for the original (black), DTFE (orange) and DTFE/CVT (blue). Central and right: following the same colour scheme, linear profiles through the filtered density field reconstruction, along the y -axis (at $x = 0.5$) and the x -axis (at $y = 0.5$).

profile (blue) appears to adhere considerably better to the original profile.

To appreciate the average trend in the density reconstruction, we filter the original, DTFE and DTFE/CVT fields with a Gaussian filter ($R_f = 0.05$). In addition to the suppressed shot noise, the resulting density level plots, shown in the central row of Fig. 8, reflect the way in which the reconstructions affect the shape of the Gaussian peak. While the DTFE reconstruction retains the shape of the original, entirely in accordance with the findings by Schaap (2007), the DTFE/CVT appears not to do so. This impression is confirmed by the two linear profiles, along the y -axis, at $x = 0.5$ and along the x -axis, at $y = 0.5$, shown in the bottom central and right-hand panels. While the DTFE reconstruction is able to follow accurately the shape of the original, the DTFE/CVT reconstruction displays considerable deviations.

We may therefore conclude that while CVT appears to suppress the shot noise effects on small scales, it is not able to follow the morphology of the mass distribution on larger scales. Since we will be filtering our fields in a similar way, this result leads us to pursue our study on the basis of the pure DTFE density estimate procedure *without* the CVT regularization.

4 INTERPOLATION METHODS

The key aspect in our investigation is the interpolation step of each of the three reconstruction methods. Here, we describe the interpolation steps of the investigated methods, the DTFE, NNFE and Kriging interpolators.

4.1 Interpolation Method I: Delaunay Tessellation Field Estimator

The most straightforward way to interpolate and/or reconstruct a density field is by linear interpolation between neighbouring data points. The linear reconstructed field is continuous throughout the sample volume. Within each interpolation interval the first derivative remains constant, although it is discontinuous at the boundaries between the intervals.

The DTFE (Schaap & van de Weygaert 2000; Schaap 2007; van de Weygaert & Schaap 2008) is the multi-dimensional equivalent of simple piecewise one-dimensional linear interpolation from an irregularly distributed set of points. DTFE generalizes the concept of natural interpolation interval to any dimension D by adopting the Delaunay tetrahedra of a multi-dimensional point set as such. It uses the adaptive and minimum triangulation properties of Delaunay tessellations to use them as adaptive spatial interpolation intervals for irregular point distributions (Bernardeau & van de Weygaert 1996).

Once the Delaunay tessellation has been constructed, and the densities at each sample point determined (see Section 3), we determine the density gradient $\nabla f|_j$ within each Delaunay tetrahedron \mathcal{T}_j from the density values $[f(\mathbf{r}_0), f(\mathbf{r}_1), f(\mathbf{r}_2), f(\mathbf{r}_3)]$ at its four vertices at locations $\mathbf{r}_0, \mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3$.

Using the density gradients in the Delaunay tetrahedra, the DTFE density value at any point $\hat{\mathbf{r}}$ can be calculated by determining in which tetrahedron it is located and subsequently computing its density estimate $\hat{f}(\hat{\mathbf{r}})$ from the linear equation,

$$\hat{f}(\hat{\mathbf{r}}) = f(\mathbf{r}_0) + \widehat{\nabla f}|_j \cdot (\hat{\mathbf{r}} - \mathbf{r}_0). \quad (7)$$

To obtain an image of the density field, one calculates these density estimates at each of the voxel locations of the image grid. For a

more detailed outline of the DTFE method we refer the readers to Appendix B.

An impression of a DTFE-interpolated field in a cosmological context is presented in Fig. 9 (top-right panel). It concerns a density field reconstruction from a data set extracted from a Millennium mock sample (top-left panel). The galaxy selection follows the distant observer approximation, i.e. following parallel lines of sight, and assumes the magnitude limit $m_r = 17.77$ of the SDSS redshift survey.

The resulting DTFE density field is the level map in the top-right panel. DTFE recovers the fine small-scale structures and at the same time adapts itself to the larger-scale structures at greater distances. It also reveals the linear interpolation artefacts, the triangular-shaped low-intensity wings. These are especially noticeable when the data points are sparse. We must note that these wings are not significant in mass, but arise when one takes a lower dimensional (one or two) section through the data.

4.2 Interpolation Method II: Natural Neighbour Field Estimator

The DTFE is a piecewise linear interpolation (C^0) method. In a sense, it is a linear version of a larger class of tessellation-based interpolation methods. Of these, Natural Neighbour interpolation (Sibson 1981; Watson 1992; Braun & Sambridge 1995) is the most well known higher order tessellation based method (for more details see van de Weygaert & Schaap 2008).

The *Natural Neighbour Interpolation* formalism is a generic higher order multi-dimensional interpolation, smoothing and modelling procedure utilizing the concept of natural neighbours to obtain locally optimized measures of system characteristics. Its theoretical basis was developed and introduced by Sibson (1981), while extensive treatments and elaborations of NN interpolation may be found in Watson (1992) and Sukumar (1998). As has been demonstrated by telling examples in geophysics (Braun & Sambridge 1995) and solid mechanics (Sukumar, Moran & Belytschko 1998; Sukumar 1998) NN methods hold tremendous potential for grid-independent analysis and computations.

According to the Sibson natural neighbour interpolation, the interpolated value $\hat{f}(\hat{\mathbf{r}})$ at a position $\hat{\mathbf{r}}$ is given by

$$\hat{f}(\hat{\mathbf{r}}) = \sum_i \lambda_{\text{NN},i}(\hat{\mathbf{r}}) f_i, \quad (8)$$

in which the summation is over the natural neighbours i of the point $\hat{\mathbf{r}}$ amongst the data points (see Fig. 5, right-hand frame). Note that a slight movement of the interpolation point will evoke a different set of natural neighbours.

The Sibson natural neighbour interpolation uses area-based (or volume in 3D) interpolation weights $\lambda_{\text{NN},i}(\hat{\mathbf{r}})$. These are determined from the volumes of the *order-2* Voronoi cells $\mathcal{V}_2(\hat{\mathbf{r}}, \mathbf{r}_j)$. To understand the concept, imagine we virtually insert the location $\hat{\mathbf{r}}$ in the spatial sample point distribution. Around its location a new Voronoi cell $\mathcal{V}(\hat{\mathbf{r}})$ is delimited (see Fig. 5, where the cell is traced by the blue edges). The virtual cell $\mathcal{V}(\hat{\mathbf{r}})$ overlaps with the original Voronoi cells \mathcal{V}_j of its natural neighbours j . The *order-2* Voronoi cells $\mathcal{V}_2(\hat{\mathbf{r}}, \mathbf{r}_j)$ are the regions of overlap, and define the region of space for which $\hat{\mathbf{r}}$ and \mathbf{r}_j are the closest ‘nuclei’.

According to the Sibson interpolation the interpolation kernel $\lambda_{\text{NN},j}(\hat{\mathbf{r}})$ is equal to the normalized area of the order-2 Voronoi cell,

$$\lambda_{\text{NN},i}(\hat{\mathbf{r}}) = \frac{\mathcal{A}_2(\hat{\mathbf{r}}, \mathbf{r}_i)}{\mathcal{A}(\hat{\mathbf{r}})}, \quad (9)$$

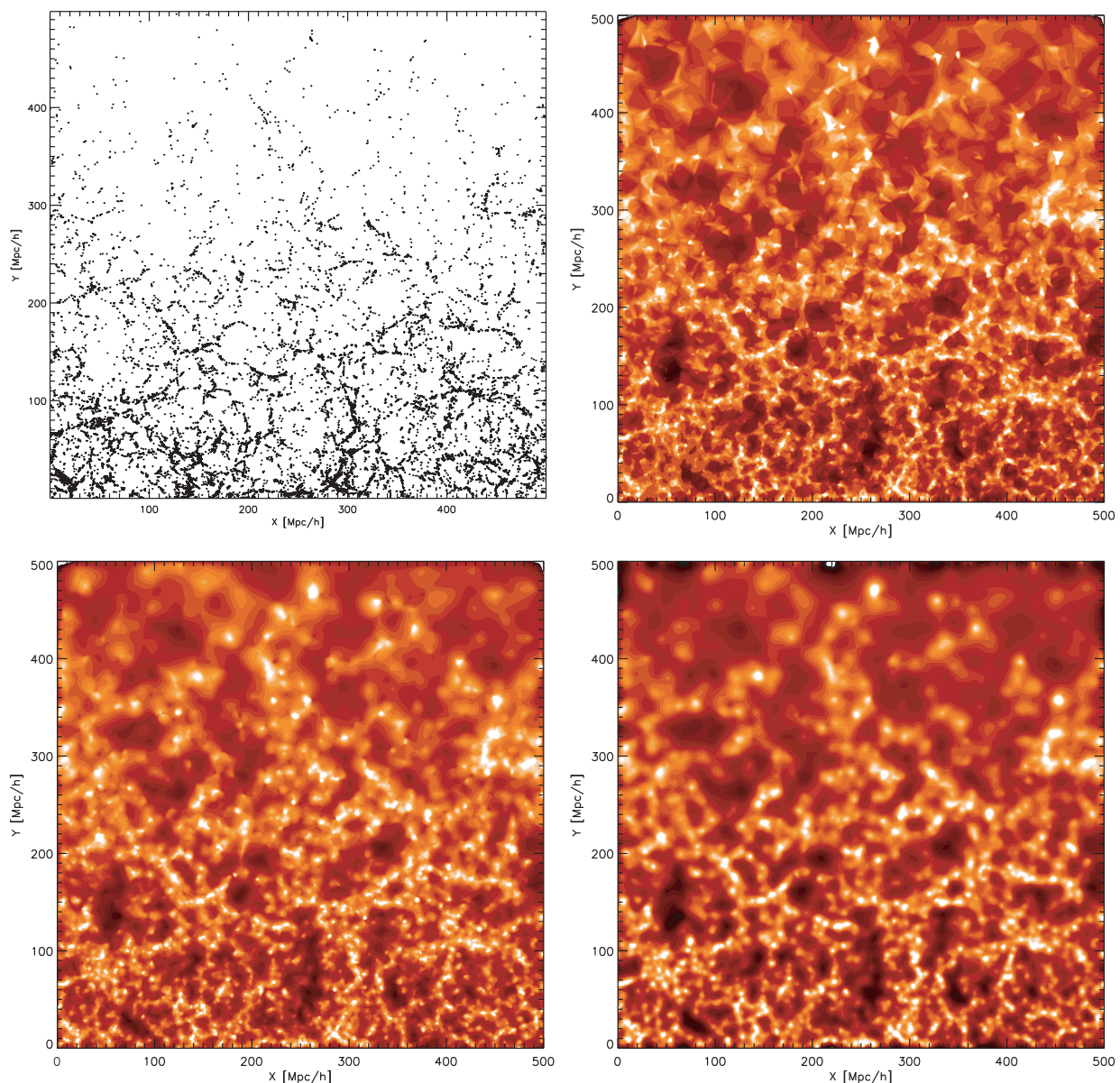


Figure 9. Density Reconstructions of SDSS mock catalogue. Top left: the Millennium galaxy mock sample, following the distant observer approximation. Top right: DTFE density field reconstruction. Bottom left: NNFE density field reconstruction. Bottom right: lognormal Kriging reconstruction.

in which $\mathcal{A}(\hat{\mathbf{r}}) = \sum_j \mathcal{A}_2(\hat{\mathbf{r}}, \mathbf{r}_j)$ is the area of the virtual Voronoi cell of point $\hat{\mathbf{r}}$ and $\mathcal{A}_2(\hat{\mathbf{r}}, \mathbf{r}_i)$ the area of the order-2 Voronoi cell $\mathcal{V}_2(\hat{\mathbf{r}}, \mathbf{r}_i)$.

Evidently, the closer one moves a point $\hat{\mathbf{r}}$ to a sample point \mathbf{r}_j , the more Voronoi cell $\mathcal{V}(\hat{\mathbf{r}})$ will overlap with the original Voronoi cell \mathcal{V}_j , and the larger the volume $\mathcal{A}_2(\hat{\mathbf{r}}, \mathbf{r}_i)$ of the order-2 Voronoi cell $\mathcal{V}_2(\hat{\mathbf{r}}, \mathbf{r}_i)$ becomes, thus increasing the weight $\lambda_{\text{NN},i}(\hat{\mathbf{r}})$. When $\hat{\mathbf{r}}$ finally coincides with one of the natural neighbours, the order-2 Voronoi cell will be identical to the old Voronoi cell at this point. The interpolated field value $\hat{f}(\hat{\mathbf{r}})$ will then be equal to the field value at that point.

Notice that the interpolation weights $\lambda_{\text{NN},i}$ are always positive and sum to 1,

$$\sum_{i=1}^N \lambda_{\text{NN},i}(\hat{\mathbf{r}}) = 1. \quad (10)$$

This property is called *partition of unity*. The resulting function $\hat{f}(\mathbf{r})$ is continuous everywhere within the convex hull of the data, and has a continuous slope everywhere except at the data themselves. At the position of the vertices the derivative of the interpolant is discontinuous.

In one-dimension DTFE and NNFE are exactly the same. When the data points are given on a regular grid, the NNFE reduces to the more familiar bi-linear (2D) or trilinear (3D) interpolation schemes. Our NNFE implementation is that of Eldering et al. (2006), a three-dimensional adaption of the two-dimensional version available in the CGAL library.

The NNFE density field reconstruction for the same Millennium mock sample as described in Section 4.1 is shown in the bottom-left panel of Fig. 9. Some of the peaks in the regions with sparse sampling appear somewhat anisotropic. This is a consequence of the discontinuous derivative at the sample point. The overall

resulting NNFE density field is well-behaved and smooth, without the artefacts that beset the DTFE reconstruction.

A drawback of the DTFE and the NNFE methods is that neither take into account the existing spatial correlations that characterize the cosmological density field. These are explicitly taken into account by the Kriging interpolation technique.

4.3 Interpolation Method III: Kriging interpolation

By basing itself on the covariance function of the density field, Kriging naturally includes the global spatial correlations of the field.

The method was named by Matheron (1963) after D. G. Krige, who started the development of the method (see Cressie 1990, for a historical overview). The interpolator has the property that it is a *best linear unbiased estimator* (Cressie 1988, 1993). Most applications of Kriging stem from the field of geostatistics, where Kriging found its origin. The applications concern measurements at irregularly scattered points which have to be translated into, for example, gold, ore or oil field reconstructions or into altitude maps.

There is a distant relationship of the Kriging interpolation to Wiener filtering techniques (Wiener 1949; Rybicki & Press 1992; Zaroubi et al. 1995; Zaroubi 2002; Erdogdu et al. 2004, 2006; Kitaura & Enßlin 2008). However, the Wiener filtering is based on a different philosophy than Kriging, in that it includes a model for the noise and is evaluated in Fourier space. The retrieved field therefore corresponds to an optimally filtered field over a range of *unknown* scales. The filter scale is dictated by the locally estimated noise, with more noise corresponding to a larger amount of smoothing. An additional disadvantage for recovering non-linear web-like features of classical Wiener filtering is that it is predicated on an underlying Gaussian distribution in the construction of the least-squares estimator for filtering the data. While advantageous for the purpose of ascertaining the exclusive presence of significant features, it therefore has the serious disadvantage of suppressing or substantially diluting non-linear structures of interest and may have difficulties in reconstructing the intricacies of the non-linear structures.

More advanced recent developments and applications of Wiener filters to the reconstruction of the density distribution have revived its potential for reconstructing the density distribution. For an extensive and in-depth overview of these developments we refer to Kitaura & Enßlin (2008) (also see Kitaura et al. 2009, 2010).

4.3.1 The Kriging formalism

Interpolation can be viewed as estimating the field value \hat{f} at location $\hat{\mathbf{r}}$ by means of a weighted linear combination of nearby known data points $f(\mathbf{r}_i)$:

$$\hat{f}(\hat{\mathbf{r}}) = \sum_{i=1}^N \lambda_i f(\mathbf{r}_i). \quad (11)$$

The main idea of Kriging, as originally formulated, is to calculate the values of weights λ_i that minimize the error with respect to the data according to the mean square variation,

$$\mathbb{E}(|\hat{f}(\mathbf{r}) - f(\mathbf{r})|^2),$$

where \mathbb{E} is the expectation over the specified quantity. We show in Jones et al. (in preparation) that this criterion can be replaced with the weaker requirement that the data and the errors be orthogonal in a statistical sense. It follows that the statistical distribution of

the field $f(\mathbf{r})$ need not be Gaussian distributed in order to achieve optimal reconstruction of the density field via Kriging.

The *Kriging equations* for the weights λ_j are

$$\sum_{j=1}^N C(\mathbf{r}_i, \mathbf{r}_j) \lambda_j = c(\mathbf{r}_i, \hat{\mathbf{r}}), \quad (12)$$

where the matrix elements of C are given by

$$C(\mathbf{r}_i, \mathbf{r}_j) = \mathbb{E}(f(\mathbf{r}_i)f(\mathbf{r}_j)), \quad (13)$$

and the vector elements $c(\mathbf{r}_i, \hat{\mathbf{r}})$ by

$$c(\mathbf{r}_i, \hat{\mathbf{r}}) = \mathbb{E}(f(\mathbf{r}_i)f(\hat{\mathbf{r}})). \quad (14)$$

From this linear system of N equations, it is straightforward to determine the N unknown weights λ_i .

While the matrix has to be inverted only once, the weights λ_i have to be specifically computed for each interpolation site $\hat{\mathbf{r}}$. After the weights have been determined, one can directly obtain the interpolated field values \hat{f} from equation (11).

4.3.2 The Kriging variogram

Usually, the covariance function $\mathbb{E}(f(\mathbf{r}_i)f(\hat{\mathbf{r}}_j))$ depends only on distance, $d = |\mathbf{r}_i - \mathbf{r}_j|$. In geostatistics, this spatial dependence of random field is usually characterized by means of a variogram $\gamma(\mathbf{r}_1, \mathbf{r}_2)$. The variogram is the mean square variation of the field values as a function of distance,

$$2\gamma(\mathbf{r}_1, \mathbf{r}_2) \equiv \mathbb{E}(|f(\mathbf{r}_1) - f(\mathbf{r}_2)|^2), \quad (15)$$

which for a stationary random field reduces to

$$2\gamma(\mathbf{h}) = \mathbb{E}(|f(\mathbf{r}) - f(\mathbf{r} + \mathbf{h})|^2). \quad (16)$$

The variogram is related to the covariance function $c(h)$,

$$\begin{aligned} c(h) &= \mathbb{E}(f(\mathbf{r}_1)f(\mathbf{r}_2)) \\ &= c(0) - \gamma(h). \end{aligned} \quad (17)$$

For practical purposes it is preferable to use a functional form for the variogram. There is a variety of such variogram models (see Cressie 1993, for a detailed description). We use the exponential expression,

$$\gamma(h) = \sigma_0 \left\{ 1 - \exp\left(-\frac{h}{h_0}\right)^p \right\}, \quad (18)$$

which represents a good fit for the variogram measured from a Millennium SDSS mock galaxy survey (see Section 2.2). To estimate the variogram for this galaxy distribution, we used the estimator (Cressie 1993),

$$\hat{\gamma}(h) = \left(\frac{1}{N_p} \sum_{i=1}^{N_p} \sqrt{|f(\mathbf{r}) - f(\mathbf{r} + \mathbf{h})|} \right)^4. \quad (19)$$

We base our estimate on a large number N_p of randomly chosen locations within the sample volume.

4.3.3 Lognormal density fields

A lognormally distributed density field f has distribution

$$P_{\text{LN}}(f) = \frac{1}{\sqrt{2\pi}S^2} \exp\left(-\frac{[\log(1+f) + S^2/2]^2}{2S^2}\right) \frac{1}{1+f}, \quad (20)$$

where S is the $S = \log(1 + \sigma)$ and $\sigma^2 = \langle f^2 \rangle$ is the variance of the density field (see Coles & Jones 1991).

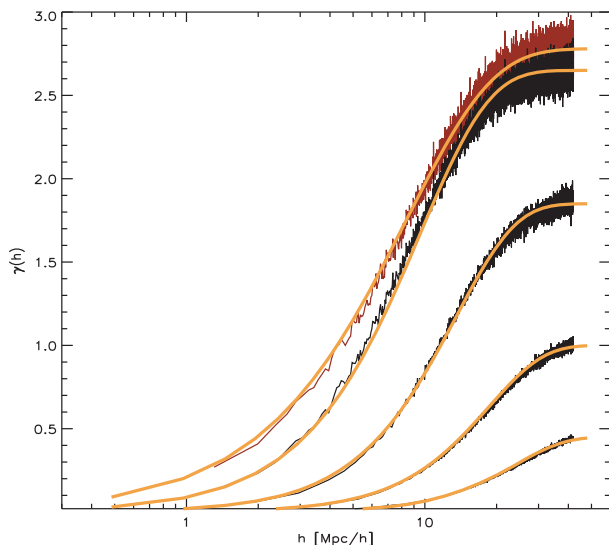


Figure 10. The measured variogram from top to bottom corresponds to the unsmoothed field (red), and the filtered fields at scales of $1 h^{-1}$ Mpc Gaussian filtered, $3 h^{-1}$ Mpc, $6 h^{-1}$ Mpc and $10 h^{-1}$ Mpc (black). The fitted variogram models according to equation (18) are shown in orange.

What might simply be referred to as the ‘Lognormal Kriging’ procedure uses the logarithm of the density field value to transform the density field data,

$$\phi(f) = \log(1 + f). \quad (21)$$

Since the density $\rho = \bar{\rho}(1 + f)$ is always positive, the application of the lognormal approach is valid everywhere and guarantees a positive definite reconstruction of the density.

The final interpolation values are obtained by taking the inverse transformation, i.e. the exponential of the interpolated data values,

$$\hat{f}(\mathbf{r}) = \exp \left\{ \sum_{i=1}^N \lambda_i \log(f(\mathbf{r}_i)) \right\}. \quad (22)$$

We will use the logarithmic value of the DTFE-interpolated field in what follows.

For such a field, Fig. 10 shows the variogram for a DTFE interpolated field (red) and for fields Gaussian smoothed on scales $R_f = 1, 3, 6, 10 h^{-1}$ Mpc. The fitted variogram model parameters (equation 18) are listed in Table 1.

In the second paper of this series on SDSS density field reconstructions, we will demonstrate that the galaxy distribution is indeed very well modelled by a lognormal distribution at scales in excess of $3 h^{-1}$ Mpc.

Table 1. Kriging Variogram parameters. Parameters obtained from simulated Millennium SDSS mock catalogue (see text).

R_f (h^{-1} Mpc)	σ_0	h_0	p	R_{\max} (h^{-1} Mpc)
0	2.78	8.4	1.2	200
1	2.65	9.6	1.5	200
3	1.85	13.5	1.72	300
6	1.0	18.6	1.9	400
10	0.45	24.0	2.1	500

4.3.4 Localized Kriging

The value of N to be used in equation (11) has so far not been defined. We chose the local neighbourhood to be the *tetrahedral natural neighbourhood*. In three dimensions this is the union of all natural neighbours of the four vertices of the Delaunay tetrahedron in which a point $\hat{\mathbf{r}}$ is located. This choice exploits the self-adaptivity to density and local shape of the Delaunay triangulation, and does not suffer the adverse effects mentioned above for the options of distance or number of neighbour selection (also cf. the discussions in Schaap 2007; van de Weygaert & Schaap 2008). Our experiments indeed confirm that the tetrahedral natural neighbourhood choice is superior to that of the two options listed above.

We found that in three dimensions the *tetrahedral natural neighbourhood* on average contains approximately 57 particles. One may extend the neighbourhood by adding a third or even more layers around it. An additional third layer would involve an average of 284 neighbours; however, the overall quality of the field reconstruction is not significantly better than with two layers despite a substantial increase in computational effort. See Jones et al. (in preparation) for more details on this.

Thus, in our key equation (22) we use the value of N that is the number of first- and second-layer vertices surrounding each point.

5 QUALITATIVE DENSITY COMPARISON

For an assessment of the performance of the three reconstruction techniques, we turn to the mock catalogues modelling the SDSS DR6 galaxy survey sample. On the basis of the knowledge of the underlying density field of the mock samples, we will be able to infer absolute statements about the quality of the reconstructions.

In this section, we will first address the visual appearance of the density field reconstructions, which will allow a qualitative and global judgement on their ability to reproduce the true density field. A quantitative error and correlation analysis will follow in the subsequent section (Section 6), while the topological properties of the reconstructions will be discussed in Section 7.

5.1 Maps of the density field

Density contour maps for the three reconstructions are shown in Figs 11 and 12. The maps concern thin slices through the density field reconstructions (and thus have zero intrinsic width).

The first set of three maps concern the reconstruction on the basis of the *full* Millennium mock galaxy sample within the boundaries of the SDSS DR6 North Galactic Cap region. The second set concerns the same region, but for the *volume-limited* mock sample.

To distinguish underdense and overdense regions, the underdense regions are indicated by means of a colour contour map while the overdense regions are represented by black contours. The contour levels in all maps are the same.

The DTFE, NNFE and Kriging maps all display the same global structure. The overall appearance of the DTFE, NNFE and Kriging density maps is strikingly similar. In the higher density regions the differences are minor, and mainly concern the coherence and anisotropy of the reconstructed filamentary (and sheet-like) features. Qualitatively speaking, the lower density regions do reveal more differences between the three methods.

As expected, the volume-limited maps fail in reproducing the small-scale structure, while they trace the overall web-like outline seen in the full sample maps.

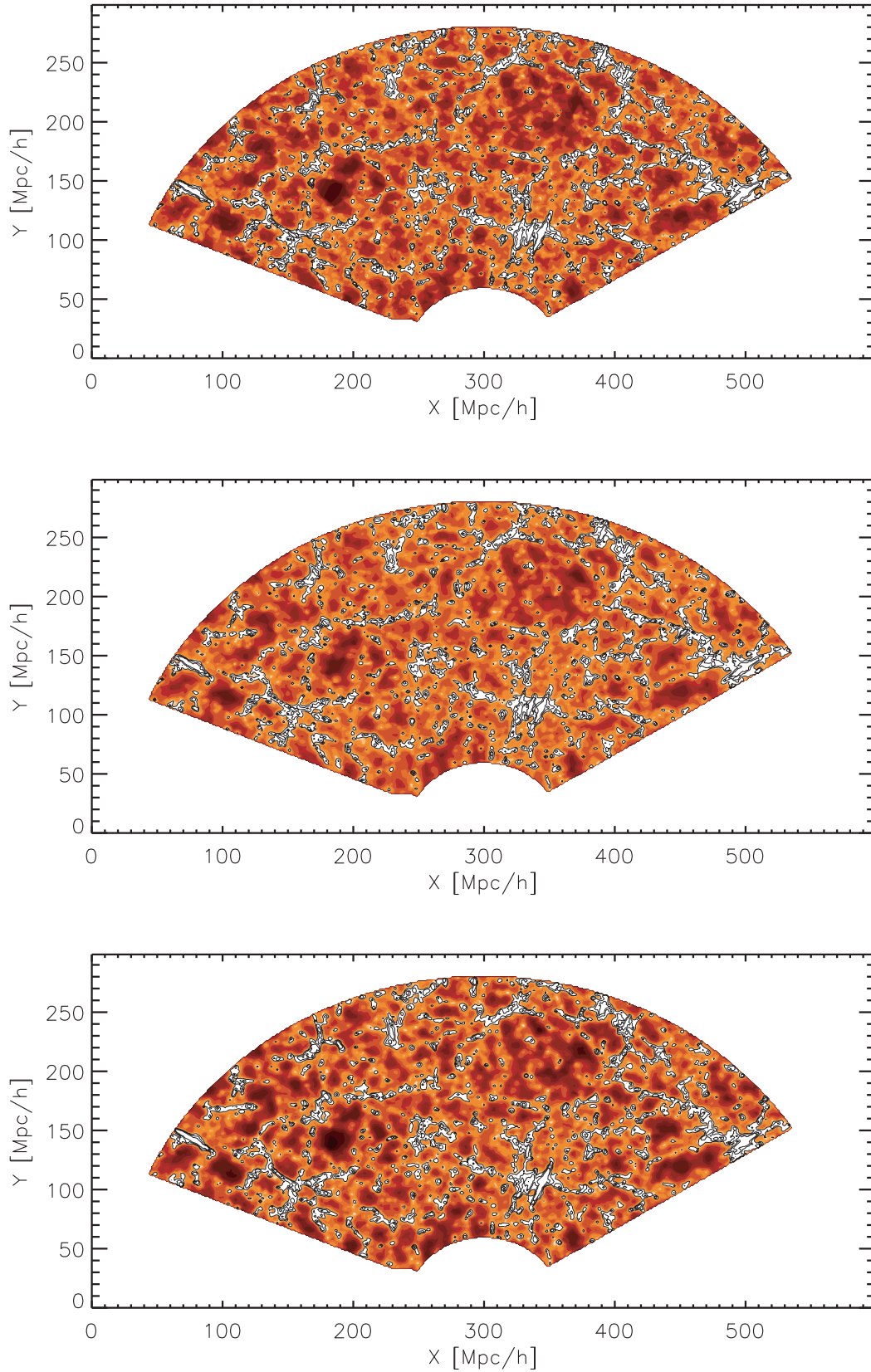


Figure 11. Comparison SDSS-DR6 reconstructions for the magnitude-limited galaxy sample. Shown are – zero width – sections through the density field reconstructions. Top: DTFE; centre: NNFE; bottom: Natural Lognormal Kriging. The coloured contour levels represent the underdense regions, at $\rho/\rho_u = [0.001, 0.002, 0.005, 0.01, 0.02, 0.03, 0.05, 0.07, 0.1, 0.2, 0.3, 0.6, 0.7, 0.8, 1]$. The white areas are the overdense regions and the black contour lines represent a density contrast $\rho/\rho_u = [1, 3, 10]$.

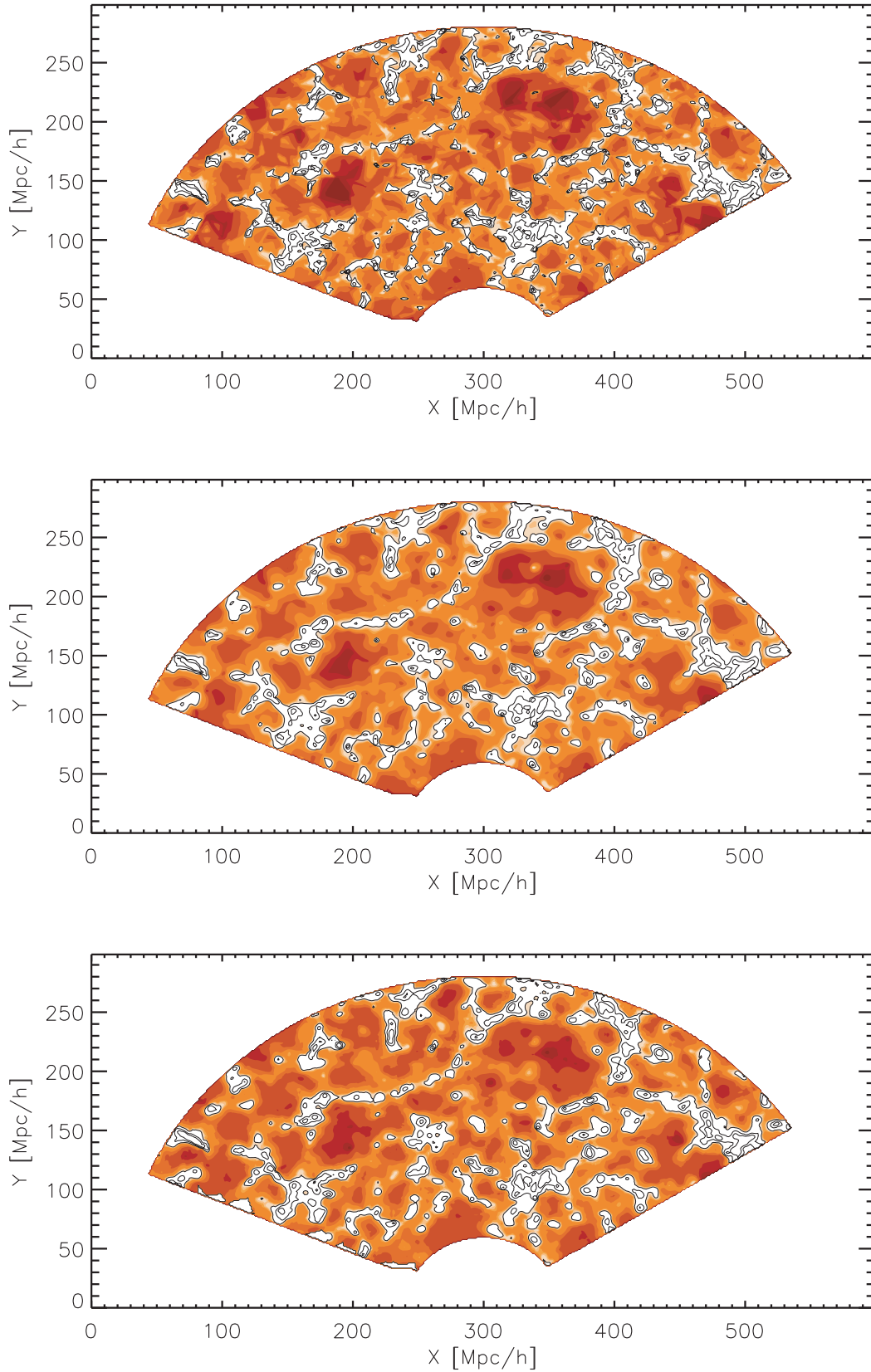


Figure 12. Comparison SDSS-DR6 reconstructions for the volume-limited galaxy sample. Shown are – zero width – sections through the density field reconstructions. Top: DTfE; centre: NNFE; bottom: Natural Lognormal Kriging. The coloured contour levels represent the underdense regions, at $\rho/\rho_u = [0.001, 0.002, 0.005, 0.01, 0.02, 0.03, 0.05, 0.07, 0.1, 0.2, 0.3, 0.6, 0.7, 0.8, 1]$. The white areas are the overdense regions and the black contour lines represent a density contrast $\rho/\rho_u = [1, 3, 10]$.

5.1.1 DTFE map

The DTFE map is remarkably accurate in outlining the tenuous web-like filamentary features, in particular in the case of the full sample reconstruction. Amongst the three reconstructions, the DTFE one looks more crispy than the NNFE and Kriging maps. It is slightly more capable in tracing the thin filamentary and sheet-like features, while one might have a slight worry with respect to the correct reproduction by NNFE and Kriging of the shape of filaments and walls. Also, we find that the DTFE maps are clearly marked by higher density contrasts, both within the overdense regions and with respect to the underdense regions.

The downside of the detailed structural reconstruction of DTFE is the more erratic nature of the DTFE contours, marked by sharp artefacts. These artefacts are particularly prominent in the field reconstruction of the volume-limited sample. They are manifestations of the linear interpolation method: when two neighbouring grid cells are located in different triangles, the field would appear to be discontinuous. The latter is mostly an impression, as sampling at finer scales would show it is just continuous. These artefacts occur in situations where the point sample density is considerably sparser than the size of the grid cells. This is also the reason why these artefacts are more prominent in the DTFE reconstruction of the volume-limited sample than in that of the full sample.

5.1.2 NNFE and Kriging

The reconstructed density maps of the two higher order schemes, NNFE and Kriging, have a considerably smoother appearance than the DTFE maps. The larger number of neighbours involved in the NNFE and Kriging reconstructions translate into the slightly more roundish contours of these structures. This also reveals itself in the absence of artefacts such as seen in the DTFE maps.

Part of the differences between the smoother higher order maps and the DTFE maps is an expression of the number of points, and field values, involved in the interpolation step (cf. e.g. equation 22). DTFE uses four points, the vertices of a Delaunay tetrahedron. NNFE involves on average 17 natural neighbours, while the Natural Kriging scheme invokes 57 neighbours.

The smoother nature of the NNFE and Kriging maps is therefore an expression of the somewhat lower information content of these filtered maps. As a result, they also have less noisy low-density regions than those seen in the DTFE maps.

5.1.3 Anisotropic structure and features

One of the crucial benefits of DTFE is that it is able to identify anisotropic features, like walls and filaments, and successfully reproduce their shape and morphology (see Schaap 2007; van de Weygaert & Schaap 2008).

From the density maps we see that NNFE and Kriging find the same filamentary structures. Overall, the impression is that DTFE and NNFE produce maps in which the Cosmic Web is more coherent than in the Kriging maps: the Kriging map mass concentrations have a slight tendency to break up more easily into clumps. This is true for both the full sample maps and the volume-limited sample maps. One exception, in the volume-limited map, seems to be the filamentary extension running from $(X, Y) = (170, 160) h^{-1} \text{ Mpc}$ to $(X, Y) = (300, 225) h^{-1} \text{ Mpc}$. One reason for the somewhat more fragmentary nature of the Kriging maps is its use of a radially symmetric covariance function, while DTFE and NNFE are based on kernels that adapt to the local shape.

5.1.4 Underdensities and voids

When turning to the underdense regions, we find that in the full density map both DTFE and Kriging delineate them at high contrast levels. The voids in the NNFE map have a lower contrast. This is partially the result of the larger NNFE neighbourhood radius in low-density areas. In this respect, the Kriging correlations assure a better performance. DTFE remains sensitive on behalf of its highly local character.

A comparison between the void population in the full sample reconstruction and that in the volume-limited sample reconstruction reveals a considerable contrast between the results for the different reconstruction methods. None of the volume-limited reconstructions contains the small voids visible in the full sample maps. This is a reflection of the absence of such depressions in the diluted point sample. Of the three reconstructions, the contrast between the two DTFE maps is less distinct than that between the two NNFE and two Kriging maps. DTFE at least manages to trace the large voids at $(X, Y) = (180, 140) h^{-1} \text{ Mpc}$, at $(X, Y) = (350, 200) h^{-1} \text{ Mpc}$ and at $(X, Y) = (460, 120) h^{-1} \text{ Mpc}$. Kriging and NNFE hardly manage to find the latter in the volume-limited maps, while the huge void complex near $(X, Y) = (350, 200) h^{-1} \text{ Mpc}$ is a largely uniform moderate underdensity in the Kriging map. Only the DTFE map reveals its true nature, a region marked by several deep voids embedded in a larger moderate underdensity.

5.2 Density profiles

To appreciate the small-scale details in the density field reconstructions, Fig. 13 displays a linear profile, along a radial distance of $500 h^{-1} \text{ Mpc}$, through the density field reconstructions. All profiles are taken along the Y -axis, at $(X, Z) = (300, 300) h^{-1} \text{ Mpc}$ (also cf. Figs 11 and 12).

The panels, from top to bottom, concern the DTFE, NNFE and Kriging reconstructions. The black lines are linear profiles through the full sample reconstructions, and the superimposed red profiles concern the volume-limited mock sample reconstructions. We also show the linear profiles through these density field reconstructions, Gaussian filtered on a scale of $R_f = 10 h^{-1} \text{ Mpc}$. The grey lines are the linear profiles through the filtered full sample density field, and the orange lines those through the filtered volume-limited sample density field.

The comparison between the linear profiles through the NNFE and Kriging reconstructions on the one hand, and the DTFE reconstruction on the other hand, leads to the following observation.

- (i) Underdense regions of NNFE and Kriging are less noisy. A nice example of this is the underdense region at $X = 200 h^{-1} \text{ Mpc}$.
- (ii) Maxima tend to be wider in the higher order schemes.
- (iii) At larger radial distance, the topology of the unfiltered reconstructions is much smoother.

In the cosmological context, the density is approximately log-normally distributed (see Coles & Jones 1991; Kofman et al. 1994; Sheth 1995; Kayo, Taruya & Suto 2001; Kitaura et al. 2010), which impels us to work the log of the inferred density field. Our adaption of the Kriging method to tessellations of lognormally distributed discrete point data is presented in detail in a separate paper (Jones et al. in preparation), where we also discuss the relationship of this procedure with the Constrained Random Field formalism developed by Bertschinger (1987) (also see Hoffman & Ribak 1991; Rybicki & Press 1992; Sheth 1995; van de Weygaert & Bertschinger 1996). We adopt the nomenclature used in the geostatistical literature.

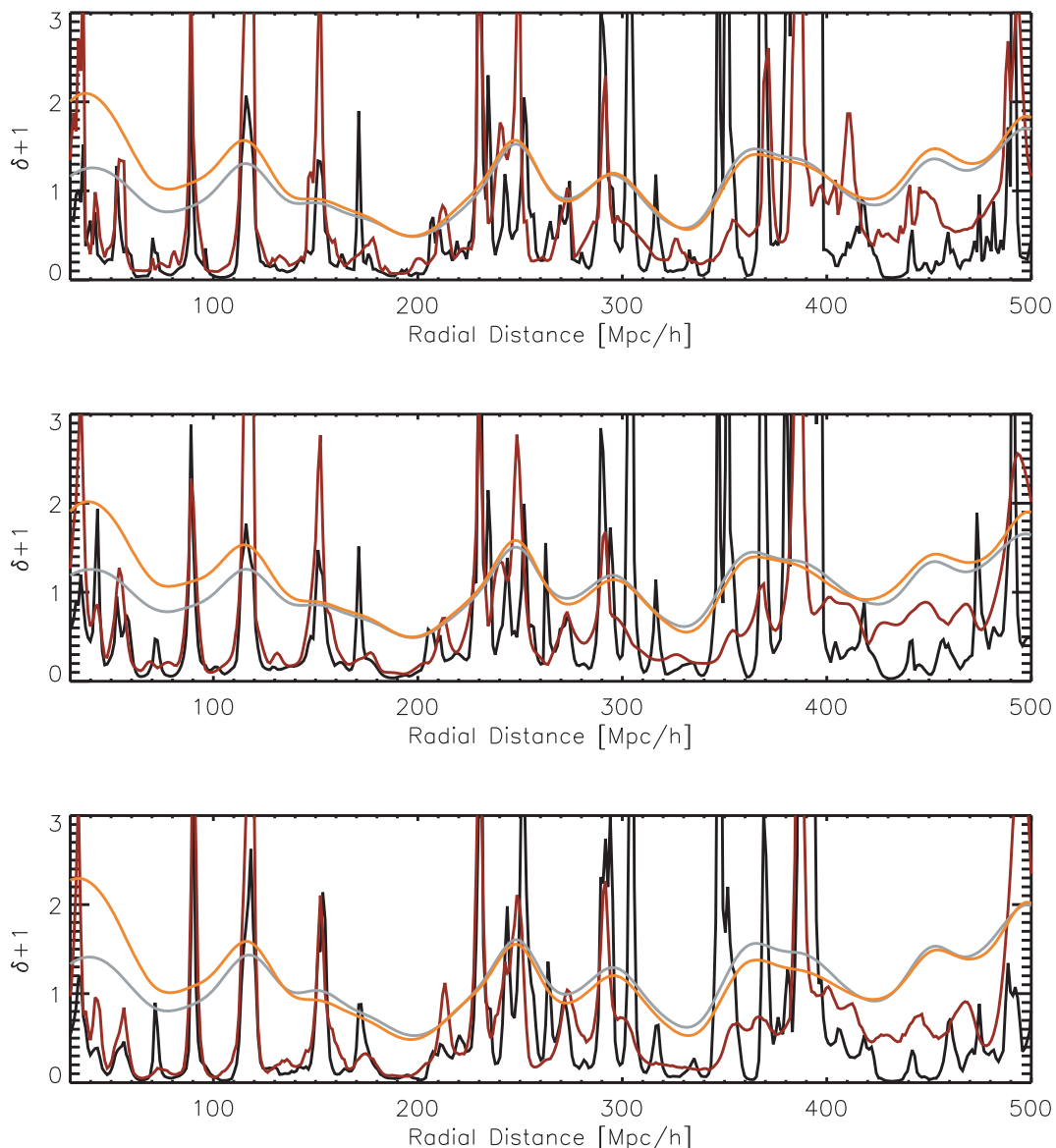


Figure 13. Linear profiles along line of sight. All profiles are taken along the Y -axis, at $(X, Z) = (300, 300) h^{-1}$ Mpc. Top: DTFE density field reconstruction; centre: NNFE density field reconstruction; bottom: Natural Lognormal Kriging reconstruction. In each panel: black line: density field reconstruction full galaxy sample; red lines: density field reconstruction volume-limited galaxy sample; grey lines: Gaussian filtered density field reconstruction full galaxy sample, $R_f = 10.0 h^{-1}$ Mpc; orange lines: Gaussian filtered density field reconstruction volume-limited sample.

6 QUANTITATIVE DENSITY FIELD ANALYSIS

For the quantitative comparison and error evaluation of the three density field reconstruction methods, we are particularly interested in the ability to recover the underlying density field from a magnitude-limited or volume-limited survey.

In this section we compare the DTFE, NNFE and Natural Lognormal Kriging reconstructions on the basis of magnitude-limited or volume-limited mock samples with that of the corresponding density fields determined from the full galaxy samples. The latter are galaxy redshift-space samples that we would hypothetically obtain if we were able to observe all galaxies in the Millennium mock sample. The magnitude- and volume-limited samples are obtained from this Millennium mock sample by imposing the observational specifications of the SDSS survey (see Section 2.1). An impression of the differences in structural resolution is provided by the visual com-

parison of the NNFE full, magnitude-limited and volume-limited sample reconstructions at the beginning of this study, in Fig. 4.

We will restrict the error analysis to the redshift-space density maps, and not assess the errors introduced by peculiar velocity distortions. These are investigated in detail in the follow-up paper.

Most of our analysis focuses on the quality of the density field reconstructions obtained from magnitude-limited samples, unless specifically stated otherwise.

6.1 Magnitude-limited survey reconstructions: correlation diagrams

The first comparison between the magnitude-limited sample density field reconstruction and the full sample density field concerns a purely local point-to-point comparison. This test involves an

inspection of the correlation diagrams of the density field value $\delta_{\text{full}}(\mathbf{r})$ of the full sample density field at location versus the corresponding density value $\delta_{\text{mock}}(\mathbf{r})$.

Since this is a strictly local comparison, and does not distinguish between systematic non-local offsets or random field fluctuations, we try to get an impression of environmental effects by simply studying a sequence of filtered density fields. Of each density field we study four Gaussian filtered versions, at filter radii of $R_f = 0.0, 1.0, 6.0$ and $10.0 h^{-1}$ Mpc. These scales represent the transition from the non-linearity ($1 h^{-1}$ Mpc) to quasi-linear and linear scales at $10.0 h^{-1}$ Mpc.

If the survey-based reconstruction were perfect, the correlation diagram should be a perfect one-to-one line. The correlation diagrams show the level of scatter, and reveal whether the reconstruc-

tion errors are dependent on density and as well expose the presence of systematic offsets.

6.1.1 DTFE correlation diagrams

Fig. 14 presents the correlation diagrams for the DTFE reconstructed fields at four filter scales, $R_f = 0.0, 1.0, 6.0, 10.0 h^{-1}$ Mpc.

Instead of a pure scatter diagram, we depict the density of the pairs $(\delta_{\text{full}}, \delta_{\text{mock}})$ by means of contour levels. The contour levels in Fig. 14 depict the number density of pairs in logarithmic bins of size $\Delta \log(1 + \delta_{\text{full}}) \times \Delta \log(1 + \delta_{\text{mock}}) = (0.02 \times 0.02)$. The levels run from one pixel per bin (black) to the maximum number density (white), in logarithmic steps of ≈ 2.5 . In each frame the black diagonal line indicates the exact one-to-one relation between

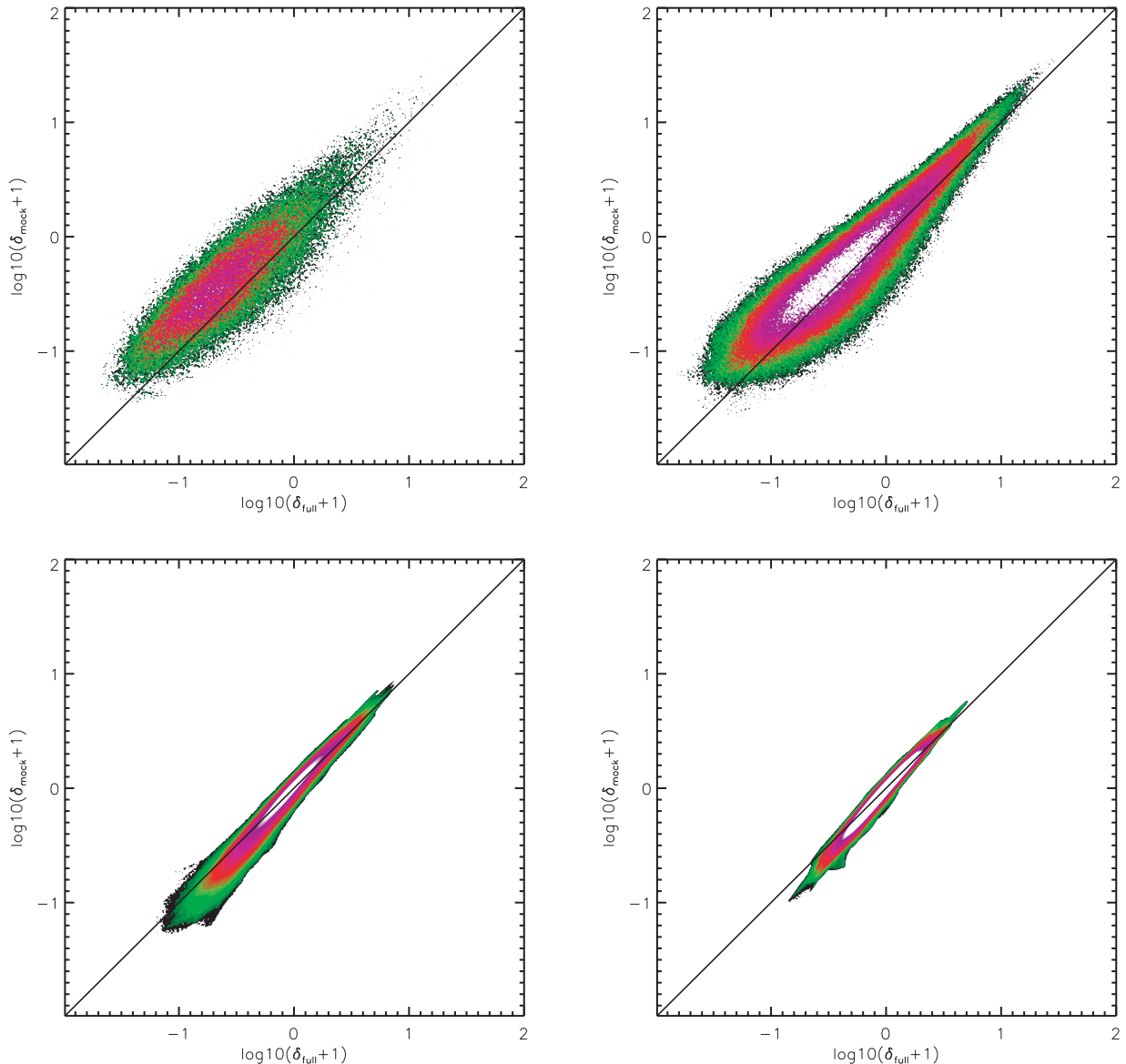


Figure 14. Correlation diagrams for DTFE density field reconstruction. Plotted are the value of the density reconstruction from the full galaxy catalogue, $\delta_{\text{full}} + 1$ (abscissa), against the density reconstruction from the magnitude-limited mock catalogues, $\delta_{\text{mock}} + 1$ (ordinate). The colours indicate the number of voxels that occupy the corresponding position in the correlation diagram, with white indicating the density pairs with highest occurrence and subsequent contour levels at logarithmic spacings of ~ 2.5 (see text). Top left: unsmoothed density field reconstructions, $R_f = 0.0 h^{-1}$ Mpc. Top right: $R_f = 1.0 h^{-1}$ Mpc. Bottom left: $R_f = 6.0 h^{-1}$ Mpc. Bottom right: $R_f = 10.0 h^{-1}$ Mpc. Note that the apparent offset of the small-scale panels is a result of cosmic variance introduced by the locally estimated density field normalization.

full sample density field and the density field following from the magnitude-limited sample.

For all filter radii, we find that over two orders of magnitude the correlation diagrams centre around a strict linear one-to-one relation. For the smaller filter radii ($R_f = 0.0 h^{-1}$ Mpc and $R_f = 1.0 h^{-1}$ Mpc) the relation appears to be slightly offset. This is a reflection of the necessarily small volumes probed by the galaxy survey at distances close to the observer, introducing cosmic variance effects through the locally estimated density normalization (see Section 6.3).

Evidently, the scatter around the linear one-to-one relation decreases as the filter radius increases. For both $R_f = 0.0 h^{-1}$ Mpc and $R_f = 1.0 h^{-1}$ Mpc, we find a more substantial scatter in the low-density regions compared to the higher density areas. The scatter over the full density range turns into a more and more uniform level as we go to larger filter radii. Indeed, for $R_f > 3.0 h^{-1}$ Mpc and $0.1 < \delta < 1$ the agreement is very good.

A related additional trend is that from a slight upturn at low-density values for $R_f = 1.0 h^{-1}$ Mpc towards a slight downturn at larger filter radii. In other words, at larger filter radii the reconstructions seem to have a systematic bias towards more underdense values.

However, we need to be careful in drawing general conclusions with respect to the low- and high-density extremes. In particular, for the large filter radii, these tend to be dominated by only a few rare objects. Because the offsets are relatively minor, we do not consider it a serious problem.

6.1.2 DTFE, NNFE and Kriging correlation diagrams

To compare the performance of the three methods, Fig. 15 shows the correlation diagrams for the density field reconstructions at a filter scale of $R_f = 3.0 h^{-1}$ Mpc. Each of the reconstructions shows an almost perfect one-to-one relation: all three methods yield unbiased reconstructions.

We cannot detect any large differences between the methods. This suggests that the remaining deviations of the reconstructed density fields are due to the initial DTFE density estimate. Again, for $R_f > 3.0 h^{-1}$ Mpc and $0.1 < \delta < 1$ the agreement is very good.

Alternative density estimators might lead to further improvements.

6.2 Magnitude-limited survey reconstructions: intrinsic smoothing scale

A characteristic of the magnitude-limited survey is the change of intrinsic spatial resolution as we proceed out to larger distances and the galaxy sample includes only the most luminous objects. While one may correct the density estimates for the accompanying dilution (see Section 3), the loss of small-scale resolution and the accompanying geometric resolution is impossible to correct. This affects any study of filaments and walls on the basis of the reconstructed density field maps.

The correction for this dilution effect is complicated by another effect, the intrinsic density-dependent resolution of the sampled galaxy surveys. Higher density regions are sampled by more galaxies, and are therefore more resolved than the low-density void regions.

We may evaluate the intrinsic resolution of the density maps in terms of the intrinsic smoothing scale R_{int} . Effectively, it is related to the characteristic galaxy separation at a given redshift and position. It is rather straightforward to incorporate the effect of the increasing dilution as a function of redshift z , on the basis of the radial selection function $\psi(z)$ (see equation 1). However, the considerable density contrasts in the inhomogeneous matter distribution render it far from trivial to find a correct expression for $R_{\text{int}}(z)$. The mean galaxy separation is biased to high-density regions and seriously underestimates the correct value $R_{\text{int}}(z)$: the intrinsic smoothing scale is smaller than average in higher density areas, and larger in low-density regions.

6.2.1 Intrinsic smoothing scale: correlation diagram

To appreciate the effect and the dependence of intrinsic smoothing length on density, we evaluate the correlation between the filtered (DTFE) density field of the full galaxy sample and the DTFE density field reconstruction on the basis of the magnitude-limited sample.

In particular, we are interested in the question in how far the *filtered* density field reconstructions are representative. We may assume they are as long as the corresponding filter scale $R_f > R_{\text{int}}(z)$. By evaluating the filtered density fields we may obtain an idea of the intrinsic smoothing scale on the basis of the density field error analysis: a sudden rise of error would indicate that $R_{\text{int}}(z) > R_f$.

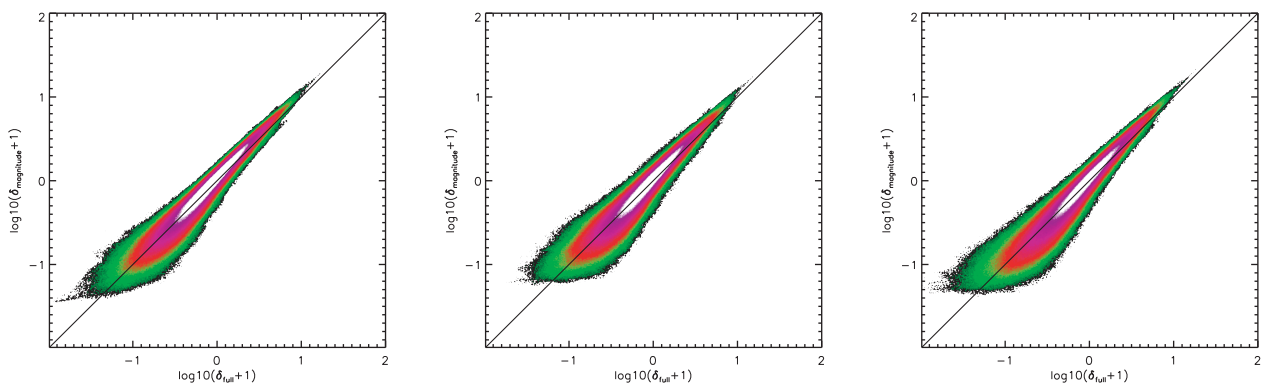


Figure 15. Correlation diagrams for DTFE, NNFE and Natural Lognormal Kriging density field reconstructions. Plotted are the value of the density reconstruction from the full galaxy catalogue, $\delta_{\text{full}} + 1$ (abscissa), against the density reconstruction from the magnitude-limited mock catalogues, $\delta_{\text{mock}} + 1$ (ordinate). The density values are those for the field filtered on a scale $R_f = 3.0 h^{-1}$ Mpc. The colours indicate the number of voxels that occupy the corresponding position in the correlation diagram, with white indicating the density pairs with highest occurrence and subsequent contour levels at logarithmic spacings of ~ 2.5 (see text). Left: DTFE reconstruction; centre: NNFE reconstruction; right: natural Lognormal Kriging.

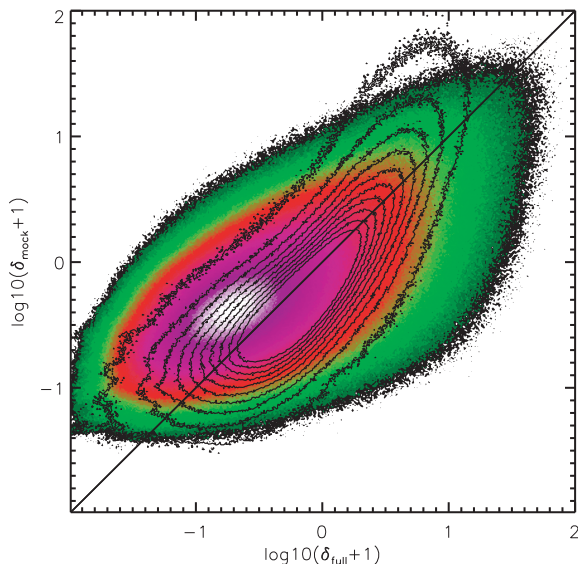


Figure 16. Intrinsic smoothing scale and density. Black contours: the correlation diagram between the $R_f = 3.0 h^{-1}$ Mpc filtered DTFE density field of the full galaxy sample (abscissa) and the unsmoothed DTFE density field reconstruction on the basis of the magnitude-limited survey (ordinate). Superimposed on the correlation diagram between the unfiltered full sample density field (abscissa) and the unsmoothed magnitude-limited mock sample density field (ordinate). The colours indicate the number of voxels that occupy the corresponding position in the correlation diagram, with white indicating the density pairs with highest occurrence and subsequent contour levels at logarithmic spacings of ~ 2.5 .

By means of the black contours in Fig. 16 we show the correlation diagram of the full sample density field filtered at $R_f = 3.0 h^{-1}$ Mpc versus the unfiltered density field on the basis of the magnitude-limited galaxy sample. It is superimposed on the (colour) correlation diagram between the unfiltered full sample density field and the unfiltered density field reconstruction from the magnitude-limited mock sample. While the latter has substantial scatter over the full density range, the correlation between the filtered field and the magnitude-limited field appears to be much tighter. It confirms the impression that reconstructed fields resemble density fields that are filtered on a particular scale. The reconstructed density maps in Figs 11 and 12 form a telling illustration of this observation.

The most outstanding feature of the filtered field correlation diagram is the curved nature of the contours, quite different from a regular linear relation. In the higher density regions we find that the densities in the magnitude-limited sample reconstruction are systematically biased to higher values. In the low to moderate density areas the trend is reverse: the densities of the raw survey density field tend to be somewhat lower than in the filtered field. This is a direct illustration of the density-dependent nature of the intrinsic smoothing length $R_{\text{int}}(z)$. Because the effective smoothing length is small in high-density regions, the low-density regions get relatively oversmoothed and so the density is relatively overestimated. More formally, it is an expression of the greater information content in the higher density regions of GRSs.

6.2.2 Filtered density field reconstructions

The filtering of a raw reconstructed density field will smooth out the high-density values. The corresponding mass is redistributed to lower density regions. We argue that this can only yield correct

density distributions if the non-linear features in the mass distribution were reproduced at the correct position and with the correct amplitude in the raw sample density field reconstruction. Information from high-density features and non-linear objects is crucial for obtaining the correct large-scale density field.

6.3 Radial error analysis

To quantify the reconstruction errors, we compare the local density values $\widehat{f}(\mathbf{r})_{\text{mock}}$ of the magnitude-limited survey reconstruction with that of the density $f(\mathbf{r})$ of the full sample. We evaluate the absolute error $\epsilon_1(\mathbf{r})$,

$$\epsilon_1(\mathbf{r}) = |\widehat{f}(\mathbf{r}) - \widehat{f}(\mathbf{r})_{\text{mock}}|, \quad (23)$$

as well as the relative error $\epsilon_2(\mathbf{r})$,

$$\epsilon_2(\mathbf{r}) = \frac{|\widehat{f}(\mathbf{r}) - \widehat{f}(\mathbf{r})_{\text{mock}}|}{\widehat{f}(\mathbf{r})}. \quad (24)$$

6.3.1 Error profiles

In Fig. 17 we plot the absolute errors $\epsilon(\mathbf{r})$ along the same radial direction as in Fig. 13. The corresponding DTFE density profile is plotted in the top panel, while the subsequent three panels depict the ϵ_1 profiles for the DTFE, NNFE and Natural Lognormal Kriging density fields.

The green lines are the error profiles of the unfiltered density field (black line in the upper panel), and the purple lines are the error profiles of the $10 h^{-1}$ Mpc smoothed field (grey line in the upper panel). We find that the errors of all three reconstruction techniques are fairly similar, with a slightly increasing trend with distance for the unfiltered density field. At a distance of approximately $100 h^{-1}$ Mpc it is of the order of unity. Beyond this distance the errors are characterized by wide peaks that are mainly due to the magnitude-limited surveys undersampling of the density field at large distances.

The errors in the $10 h^{-1}$ Mpc filtered field remain small over the whole reach of the linear profile, averaging an error level of around 10 per cent. At intermediate distances, $200 h^{-1} \text{ Mpc} < R < 400 h^{-1} \text{ Mpc}$, the DTFE errors are somewhat lower than those of the other methods. For all three methods, the largest errors are found at close distances. This is attributed to the survey mask, which is relatively thin in our immediate cosmic environment, i.e. at distances $R < 100 h^{-1}$ Mpc. Note that the filtered field errors at large distances appear to converge to the error profile of the unfiltered field reconstruction. It is a reflection of the fact that the intrinsic smoothing length becomes comparable to the filter radius.

6.3.2 Radially averaged error profiles

By averaging the errors of the magnitude-limited survey density field reconstructions over radial shells we obtain an idea of error trends as a function of distance. To minimize edge effects we exclude locations within 15 voxels from the edge of the survey volume.

The radially averaged error profiles for the DTFE density field reconstruction are shown in the top panel in Fig. 18. It concerns the absolute error ϵ_1 (solid line) and relative error ϵ_2 (dashed line) for four different filtered DTFE density fields. These are Gaussian filtered fields at filter scales of $R_f = 1.0, 3.0, 6.0$ and $10.0 h^{-1}$ Mpc.

At nearby distances $R < 100 h^{-1}$ Mpc the average error trend is that of decreasing absolute and relative errors. This is a reflection of

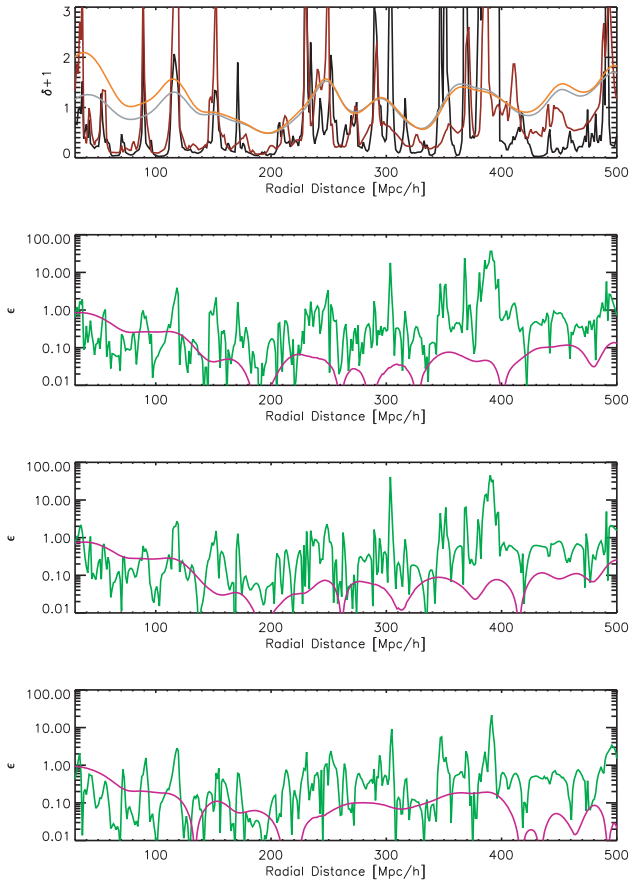


Figure 17. Linear error profiles. The absolute error ϵ_1 (equation 23) as a function of distance, along the same axis as the density profiles in Fig. 13. Upper panel: the density profile through the DTFE density field (see Fig. 13 for description). The subsequent frames are the error profiles for the three reconstruction methods. The green line represents the error of the unsmoothed density field, and the purple line that of the $10.0 h^{-1}$ Mpc filtered field. Second frame from top: DTFE density field; third frame from top: NNFE density field; bottom: Natural Lognormal Kriging density field.

the small and unrepresentative survey volumes involved. The fact that this effect appears to be most prominent for the largest filter scale, $R_f = 10.0 h^{-1}$ Mpc, is indicative. At this scale the reconstruction involves only a few independent wavevectors that constitute the resulting density field. Errors are enhanced by the way in which the weight factors $w(z)$ (equation 3) are determined. We do this by fitting the selection function on the basis of the data (see Section 2.4). At close distances this fit is heavily influenced by the presence or absence of large superstructures, which easily evokes systematic offsets in the local density estimate. It would probably be better to deal with such systematics on the basis of volume-limited samples at close distances $R < 150 h^{-1}$ Mpc.

6.3.3 DTFE, NNFE and Kriging error profiles

The bottom panel of Fig. 18 compares the relative error trend for the three different reconstruction formalisms. The radial error profiles determined are shown for the four filter scales $R_f = 1.0, .0, 6.0$ and $10.0 h^{-1}$ Mpc. The DTFE error profiles are marked by the solid lines, the NNFE ones by the dashed lines and the Lognormal Kriging ones by the dot-dashed lines.

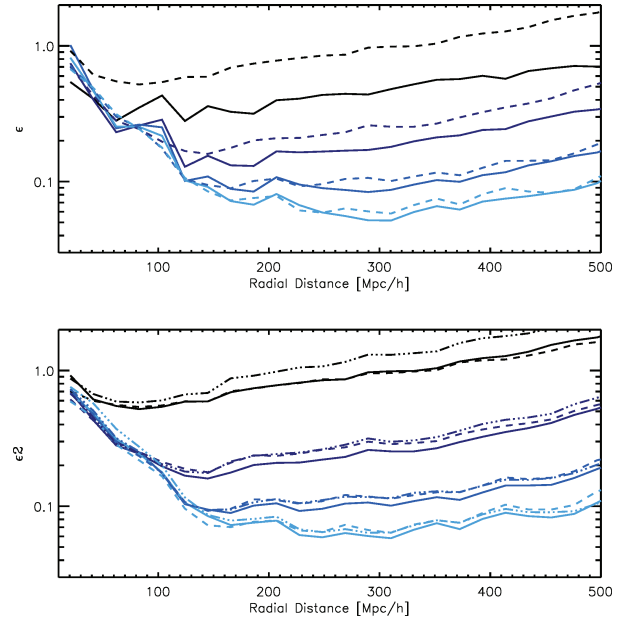


Figure 18. Radially averaged error profiles. Top: the absolute error profile $\epsilon_1(r)$ (solid line) and relative error $\epsilon_2(r)$ (dashed line) for the magnitude-limited survey DTFE density field reconstruction. Going from dark to light blue, the profiles represent the error profiles through a sequence of filtered density fields: $R_f = 1.0 h^{-1}$ Mpc (black), $R_f = 3.0 h^{-1}$ Mpc, $R_f = 6.0 h^{-1}$ Mpc and $R_f = 10.0 h^{-1}$ Mpc (light blue). Bottom: the radially averaged relative error profile $\epsilon_2(r)$, for the DTFE density field reconstruction (solid lines), the NNFE density field reconstruction (dashed lines) and the Natural Lognormal Kriging reconstruction (dot-dashed lines). The colours correspond to the same Gaussian filtered fields as specified for the top panel.

For all three methods the errors at the smallest filter scale, $R_f = 1.0 h^{-1}$ Mpc, are substantial, hardly ever better than 50 per cent. The errors decrease rapidly with filter scale, such that at scales $R_f > 6 h^{-1}$ Mpc the density field can be reconstructed with reasonable accuracy. Beyond a distance of $100 h^{-1}$ Mpc, the relative errors throughout the whole survey volume do not exceed the 10 per cent level.

Overall, we find strikingly small differences between the three methods. None performs distinctly better than any of the others. On a more detailed level, we may observe two differences. First, we see that the Kriging errors at small scales are relatively high. Secondly, DTFE appears to perform somewhat better at nearly all scales, except the smallest scale $R_f = 1.0 h^{-1}$ Mpc. In terms of density errors, the relatively simple and direct DTFE method seems to work best.

6.4 Volume-limited survey density field reconstructions

A volume-limited survey circumvents the resolution complications encountered in magnitude-limited surveys. Instead of having to correct for the steadily decreasing resolution at larger distances, volume-limited surveys have the advantage of a uniform sample resolution. This renders the geometric and topological analysis of structures considerably more straightforward. Moreover, in addition to removing the problem of varying spatial resolution, the major advantage of using volume-limited samples is that one is using the same type and luminosity of galaxies at different distance. The major disadvantage of volume-limited galaxy samples is their relatively low resolution, resulting from the necessity to uniformly sample galaxies throughout the sample volume. In essence, it involves a

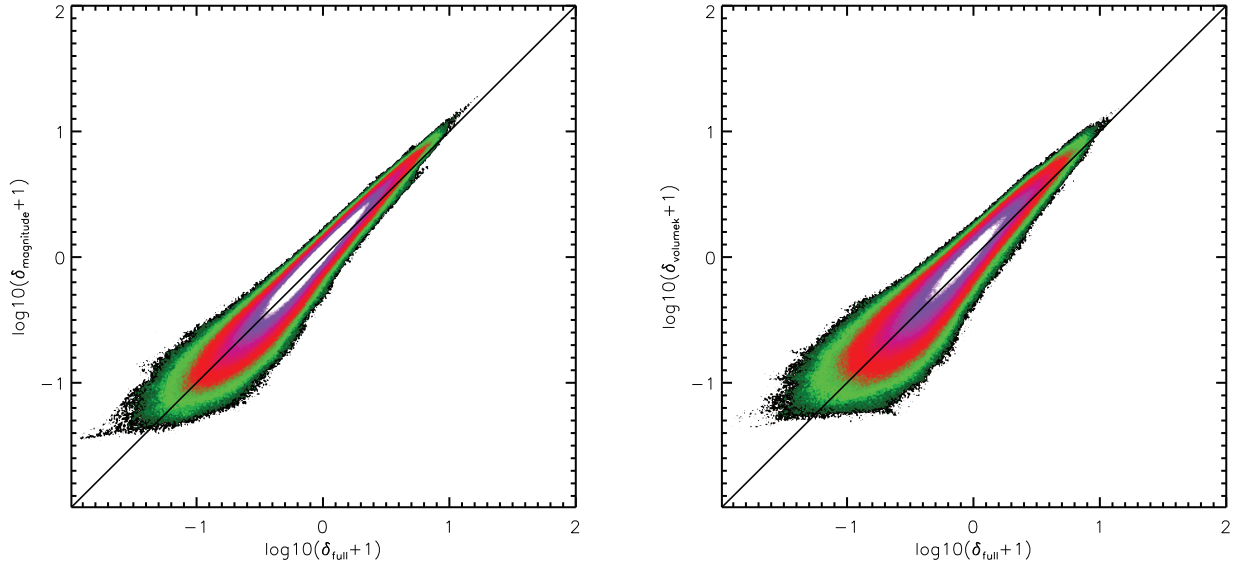


Figure 19. Density field correlation diagrams. Left: correlation diagram between full sample DTFE density field reconstruction (abscissa) and the DTFE density field reconstruction on the basis of the magnitude-limited sample (ordinate). Right: correlation diagram between full sample DTFE density field reconstruction (abscissa) and the DTFE density field reconstruction on the basis of the volume-limited sample (ordinate). All fields are Gaussian filtered on a scale $R_f = 3.0 h^{-1}$ Mpc.

trade-off between spatial resolution and a sufficiently large and representative sample volume.

To test the performance of density field reconstructions on the basis of a volume-limited GRS, we take a volume-limited sample from the DeLucia mock sample based on the Millennium simulation (De Lucia & Blaizot 2007), restricted to a volume in between redshifts $z = 0.02$ and 0.1 and containing galaxies brighter than $M_r = -20.45$.

The resulting galaxy sample can be seen in the bottom panel of Fig. 4 (along with the corresponding NNFE density field contours). With respect to the magnitude-limited sample (central panel), let alone the full sample, it clearly lacks spatial resolution. This may also be appreciated from the three corresponding density field reconstructions in Fig. 12. The loss of small-scale details is obvious: close inspection and comparison with the full sample reveals the loss of fine filamentary features separating small voids. They either merge into larger surrounding overdensities or get lost in enlarged void regions.

6.4.1 Density field correlation diagrams

Fig. 19 compares the correlation diagrams for the DTFE density field reconstructions on the basis of magnitude (left) and volume-limited samples (right). Overall, they occur rather similar, although there are some important details in which they differ. A particular one concerns the more extended maximum of the correlation diagram in the case of the magnitude-limited sample density field. In other words, the errors of the magnitude-limited sample density field are usually smaller than those for the volume-limited sample density field.

6.4.2 Radially averaged error profiles

Fig. 20 compares the profiles of the radially averaged absolute errors $\epsilon_1(r)$ (equation 23), and of the radially averaged relative errors $\epsilon_2(r)$ (equation 24) for the DTFE density field reconstructions on the basis

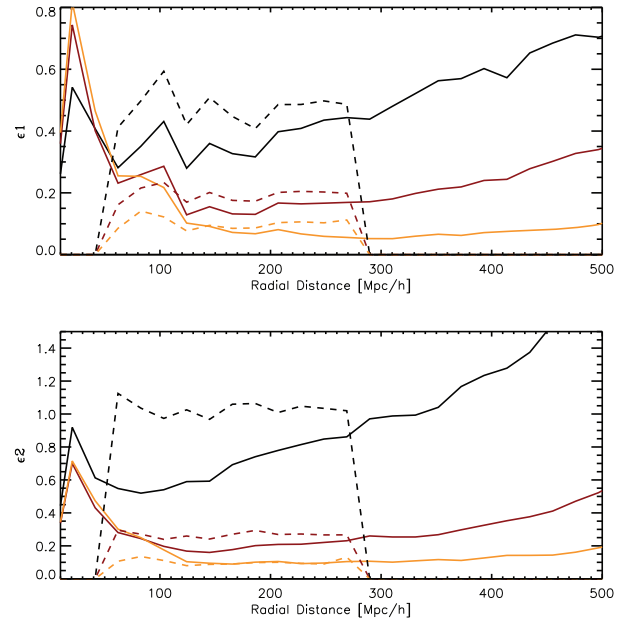


Figure 20. Radially averaged error profiles. Top: radially averaged profile of absolute error $\epsilon_1(r)$. Bottom: radially averaged profile of relative error $\epsilon_2(r)$. Solid lines, both panels: the error of the DTFE density field reconstruction on the basis of the magnitude-limited sample. Dashed lines, both panels: the error of the DTFE density field reconstruction on the basis of the volume-limited sample (within the limits of the volume-limited sample volume). Each panel shows the error profiles for three different filter scales: $R_f = 1.0$ (orange), 3.0 (red) and $10.0 h^{-1}$ Mpc (black).

of the magnitude-limited (solid lines) and on the basis of the volume-limited surveys (dashed lines). The error profiles are assessed for three Gaussian filter radii, $R_f = 1.0, 3.0$ and $10.0 h^{-1}$ Mpc.

The first observation from Fig. 20 is that the errors of the volume-limited sample are uniformly distributed throughout the survey volume, as might be expected for a statistically uniform sample.

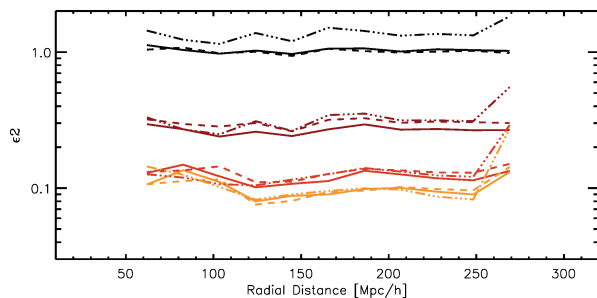


Figure 21. Radially averaged error profiles for DTFE, NNFE and Kriging density field reconstructions on the basis of volume-limited galaxy sample. The relative error $\epsilon_2(r)$ is plotted for three filter scales: $R_f = 1.0$ (orange), 3.0 (red) and $10.0 h^{-1}$ Mpc (black). Solid lines: DTFE error profiles. Dashed lines: NNFE error profiles. Dot-dashed lines: Natural Lognormal Kriging profiles.

We also find that the errors of the magnitude-limited sample density field reconstructions are systematically lower than those for the volume-limited sample. This remains so up to the edge of the volume-limited sample, at $R \approx 300 h^{-1}$ Mpc, where the sampling density of the volume-limited and magnitude-limited sample is equal.

6.4.3 DTFE, NNFE and Kriging error profiles

When comparing the error profiles for the three different density field reconstructions we hardly find any significant differences. This is confirmed by Fig. 21, which present the radially averaged relative error profiles $\epsilon_2(r)$ for the DTFE, NNFE and Kriging reconstructions at three different filter radii.

The uniformity of the errors appears to suggest that a major source for the observed errors has to be found in the density estimate itself, rather than in the interpolation technique. If there are any differences, we may argue that these concern a slightly better performance of DTFE.

7 TOPOLOGICAL ANALYSIS

To probe the global pattern of the mass distribution we turn to the topological structure of the SDSS survey. This is largely dependent on the higher order correlations in the density field. The error analysis described in the previous sections would not necessarily be able to detect key differences in the large-scale topology. In this section we will seek to evaluate the quality of the topological renderings of the density field, where we focus on the structure defined by the voids in the cosmic density field.

There are various approaches to studying the topological structure of the large-scale mass distribution. One option for characterizing the global topology of the cosmic matter distribution is in terms of four Minkowski functionals (Mecke, Buchert & Wagner 1994; Schmalzing et al. 1999). These are solidly based on the theory of spatial statistics and also have the great advantage of being known analytically in the case of Gaussian random fields. In particular, the *genus* of the density field has received substantial attention as a strongly discriminating factor between intrinsically different spatial patterns (Gott, Dickinson & Melott 1986; Gott et al. 1989; Park et al. 1992; Hoyle, Vogeley & Gott 2002; Gott et al. 2008; Zhang, Springel & Yang 2009). An attempt to extend the scope of the Minkowski functionals towards locally defined topological measures of the density field has been developed in the SURFGEN

project defined by Sahni and Shandarin and their coworkers (Sahni, Sathyaprakash & Shandarin 1998; Shandarin, Sheth & Sahni 2004). The main problem for these formalisms remains the user-defined, and thus potentially biased, nature of the continuous density field inferred from the sample of discrete objects.

Here, we specifically address the topology of the SDSS galaxy distribution on the basis of the void population. To this end, we segment the galaxy distribution into void patches by means of the WVF (Platen et al. 2007). We test the watershed segmentation of the DTFE density field obtained from the magnitude-limited mock sample by comparing it to the watershed segmentation of the full galaxy sample density field. The topological errors are quantified according to the watershed segmentation of both fields.

7.1 Watershed Void Finder

The WVF is an implementation of the *Watershed Transform* for segmentation of images of the galaxy and matter distribution into distinct regions and objects and the subsequent identification of voids (Platen et al. 2007).

The basic idea behind the watershed transform finds its origin in geophysics. It delineates the boundaries of the separate domains, the *basins*, into which yields of e.g. rainfall will collect. The analogy with the cosmological context is straightforward: *voids* are to be identified with the *basins*, while the *filaments* and *walls* of the Cosmic Web are the ridges separating the voids from each other.

With respect to the other void finders the watershed algorithm has several advantages. Because it identifies a void segment on the basis of the crests in a density field surrounding a density minimum it is able to trace the void boundary even though it has a distorted and twisted shape. Also, because the contours around well-chosen minima are by definition closed the transform is not sensitive to local protrusions between two adjacent voids. The main advantage of the WVF is that for an ideally smoothed density field it is able to find voids in an entirely parameter free fashion.

The WVF consists of eight steps, which are extensively outlined in Platen et al. (2007). For the success of the WVF it is of utmost importance that the density field retains its morphological character. To this end, the two essential first steps relate directly to DTFE, which guarantees the correct representation of the hierarchical nature, the web-like morphology dominated by filaments and walls, and the presence of voids (van de Weygaert & Schaap 2008). Because in and around low-density void regions the raw density field is characterized by a considerable level of noise, a second essential step suppresses the noise by an adaptive smoothing algorithm which in a consecutive sequence of repetitive steps determines the median of densities within the *contiguous Voronoi cell* surrounding a point. The determination of the median density of the natural neighbours turns out to define a stable and asymptotically converging smooth density field fit for a proper watershed segmentation. The subsequent central step of the WVF formalism consists of the application of the discrete watershed transform on this adaptively filtered density field.

A related tessellation-based method for void identification, ZOBOV (Neyrinck 2008), does yield similar results as WVF (see Colberg et al. 2008). It demonstrates the successful application of tessellation-based techniques to identify structures within the cosmic matter distribution. In addition to the WVF and ZOBOV there is an array of void identification procedures (see e.g. Kauffmann & Fairall 1991; El-Ad & Piran 1997; Hoyle et al. 2002; Arbabi-Bidgoli & Müller 2002; Plionis & Basilakos 2002; Patiri et al. 2006; Colberg et al. 2005, 2008; Shandarin et al. 2006). The ‘voidfinder’ algorithm

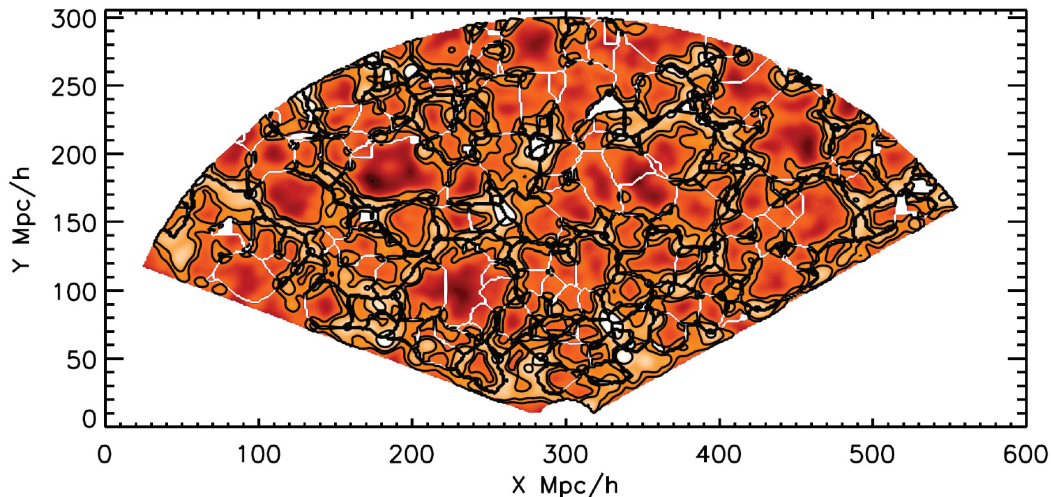


Figure 22. Watershed segmentation. The WVF watershed boundaries (white – and black edged) inferred from the Millennium magnitude-limited SDSS mock galaxy sample. They are superimposed on the density field of the full SDSS mock galaxy sample (coloured), and the contours for the overdense regions in the magnitude-limited survey density field, filtered on a $R_f = 2.0 h^{-1}$ Mpc scale.

of El-Ad & Piran (1997) has been at the basis of most voidfinding methods. However, this successful approach is not able to analyse complex spatial configurations in which voids may have arbitrary shapes and contain a range and variety of substructures, which lies at the heart of our analysis.

7.2 WVF void population maps

Fig. 22 shows the watershed segmentation generated by the mock catalogue. It is marked by the watershed boundaries superimposed on the contour maps of the density field of the full galaxy sample. These boundaries are visible as the white or thick black cell edges (dependent on the width of the boundary). The softer contour levels indicate the overdense regions in the $R_f = 2.0 h^{-1}$ Mpc filtered full sample density field.

For an impression of the ability to infer the correct watershed segmentation – and void population – from the magnitude-limited survey, we compare it with the segmentation for the full galaxy sample. In Fig. 23, the watershed boundaries of the latter are marked by red edges, while the ones for the magnitude-limited sample are marked by the black lines.

At distances up to $R \sim 200 h^{-1}$ Mpc there is overall a reasonable agreement between the two void segmentations. Beyond that distance, we find that the larger segments of the magnitude-limited sample – a consequence of the decreasing structural resolution of the survey – encompass several smaller void segments from the full sample. Beyond that distance we also find some regions with strong differences between the two segmentations. The segmentation within the large voids in the two zoom-ins illustrates this. In the magnitude-limited survey segmentation, the weaker bridges between the smaller voids visible in the full sample have vanished.

It is a clear illustration of the fact that at large distances the only topological information retained in the magnitude-limited density field is the skeleton defined by the strongest and the most overdense filaments and walls, locations traced by the brightest galaxies. The more tenuous filigree of smaller filaments within the low density regions is lost.

7.3 Topological error definition: false-splits and false-mergers

We evaluate the performance of the magnitude-limited survey watershed segmentation by comparing its void patches with those of the full galaxy sample.

As described extensively in Platen et al. (2007), the errors can be classified, to the first order, by *false splits* and *false mergers*. A *false split* is the situation in which a void segment from the reference field splits into two or more void patches. The reverse situation is that of a *false merger*, where two void segments in the reference field merge into one void patch in the segmentation of the magnitude-limited survey.

False split measure

We define a measure for the false split errors. Starting from the void segmentation in the full sample reconstruction, we find for each void segment the overlapping void patches in the magnitude-limited segmentation. The relative fraction of overlap of these survey void patches with the original void segment is used as a measure of significance of the overlap. If there are two or more significant overlaps, we classify the configuration as a false split.

Fig. 24 illustrates the above. The large black circle represents a void patch in the original segmentation, while the four grey circles are void segments in the survey segmentation. The grey shaded areas belong to the areas of the three reconstructed patches j that intersect the black circle, while the resulting dashed areas represent the intersection of the original with each of these void patches. If the surface area of circle j is equal to \mathcal{A}_j , and that of the overlap with the black original segment $\mathcal{A}_{j \cap 0}$, then the false split measure

$$f_j^{\text{FS}} = \frac{\mathcal{A}_{j \cap 0}}{\mathcal{A}_j} \quad (25)$$

decides on whether this is a significant overlap or not. We deem this to be so if $f_j^{\text{FS}} \geq 0.6$. In Fig. 24, void1 and void2 would correspond to significant overlaps. Void3 would be excluded as such and would only correspond to a slight shift of the boundary. Because of the

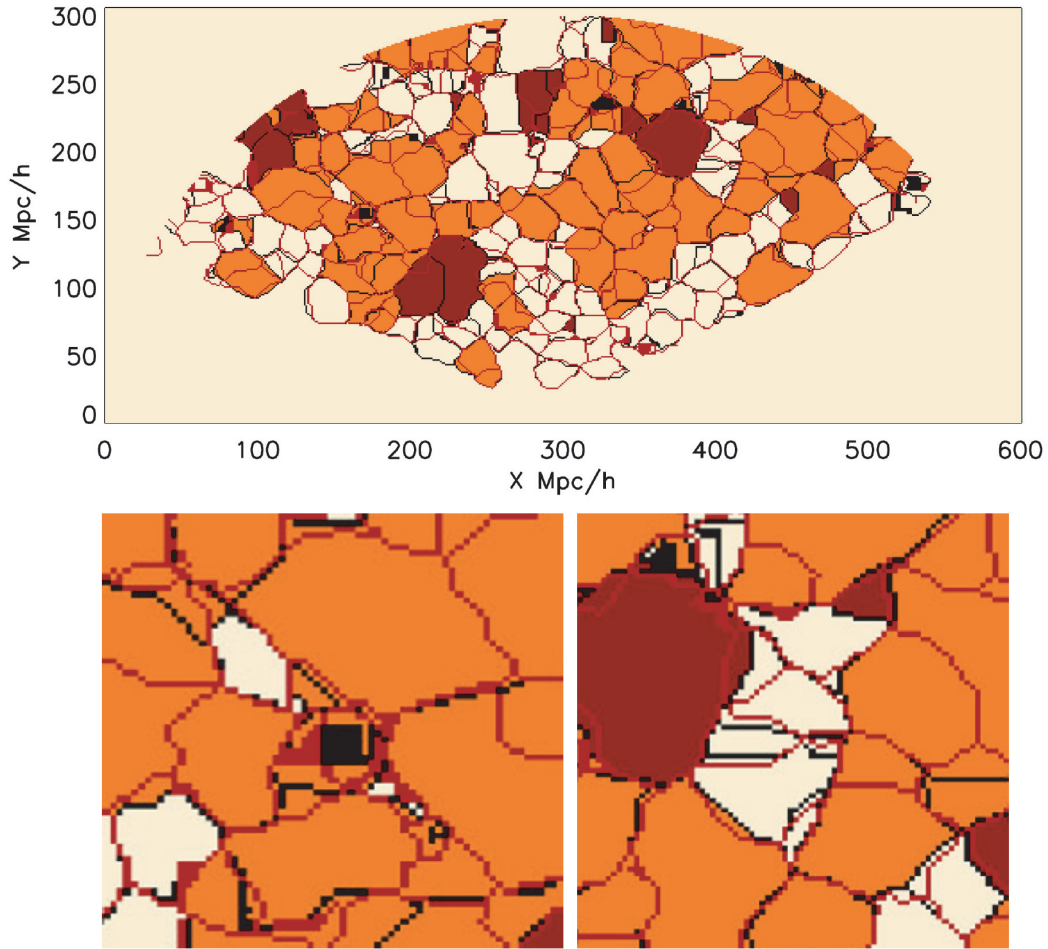


Figure 23. WVF segmentation comparison. The WVF segmentation boundaries of the full SDSS mock galaxy sample are indicated by red, and the ones for the magnitude-limited sample by black lines. The colour of the segments indicates the nature of the topological errors. Orange: *false mergers*; red: *false splits*. Below: the zoom-ins show two regions with a serious mismatch between the WVF segmentations of full and magnitude-limited sample.

two significantly overlapping voids, this configuration would be classified as a false split.

False merger measure

An almost equivalent measure can be defined for a false merger. Since by symmetry, a false merger can be considered a false split in the full (original) sample segmentation, we may simply reverse the definition of the false split measure. If l is a void patch in the original field, with area/volume \mathcal{A}_l , and its relative overlap with a large void patch in the magnitude-limited survey is

$$f_l^{\text{FM}} = \frac{\mathcal{A}_{l \cap o}}{\mathcal{A}_l}, \quad (26)$$

we deem it a significant overlap when $f_l^{\text{FM}} > 0.6$. When there are at least two original void segments that overlap significantly with one in the magnitude-limited survey, we may consider it a false merger.

Correctly identified voids

Having defined the false split and false merger errors, we may specify the meaning of a *correctly identified* patch. A *correct* void patch in the magnitude-limited survey is a void segment which overlaps for at least 60 per cent with a void segment in the original field, as well as the other way around. These two conditions prevent a void from being either a false split or a false merger.

7.4 Spatial distribution topological errors

Fig. 23 shows the spatial distribution of the topological errors. The image contains the watershed segmentation for the full galaxy sample, indicated by the black solid lines. These are superimposed on the watershed segmentation for the magnitude-limited survey, indicated by the red solid lines.

The false mergers are indicated by orange patches. The dark red patches represent the false splits, while the correctly reproduced patches are represented as white cells. It is directly apparent that false mergers are far more abundant than false splits.

A visual comparison between the two segmentations (also cf. Fig. 22) reveals the disappearance of void boundaries seen in the original full sample matter distribution. They disappear within the large void regions found in the magnitude-limited survey. In other words, these minima are absorbed by one large encompassing void. It is a result of the tenuous and usually underdense nature of the walls in these regions, so that only a few galaxies have to fall out of the survey to evoke a merging of voids.

None the less, the coherence of the large persistent watershed lines remains strong throughout the volume. The defining skeleton of the cosmic mass distribution remains largely intact in a magnitude-limited survey.

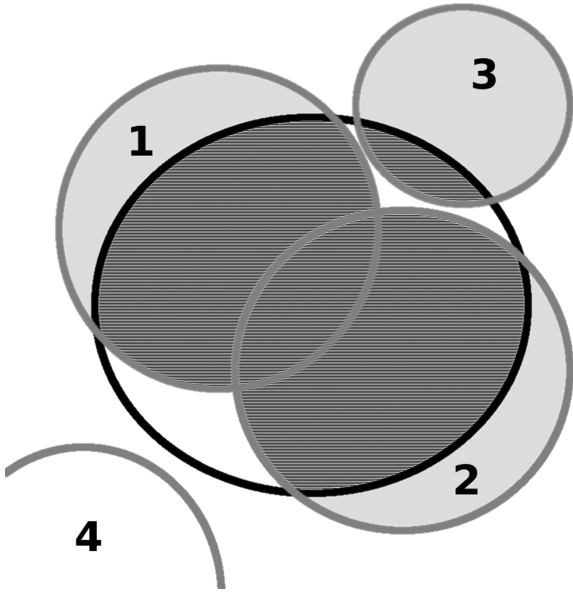


Figure 24. Definition of the false split measure. The four grey circles represent the patches from the reconstructed segmentation, the black circles would refer to a patch in the original segmentation. The grey shaded areas \mathcal{A}_j belong to the areas of the three reconstructed patches that have an intersection with the black circle. The dashed areas represent intersections between the original and the reconstructed segmentation $\mathcal{A}_j \cap O$.

7.5 Topological error characteristics

The quality of the void representation of the survey can be inferred from the percentage of correctly identified void patches, as well as from the percentage of topological errors, i.e. the total number of false splits and false mergers. Here, we assess the correct identifications and topological errors as a function of distance R . We accomplish this by counting the number of correctly identified void patches in radial shells, along with the number of error patches in the same shells.

Note that the percentage of correct and of incorrect void identification does not always add up to 100 per cent, since topological errors may be far more complex than just a false split and/or false merger. This happens with multiple additions or disappearances of void-walls.

Fig. 25 plots the identification percentages as a function of radial distance R , for all three reconstruction techniques (DTFE: blue; NNFE: orange; Kriging: grey). The top panel concerns the $3 h^{-1}$ Mpc filtered field, and the bottom panel the $10 h^{-1}$ Mpc filtered field. In the case of the $3 h^{-1}$ Mpc field we see a gradual and continuous decrease of correct void identifications as we move outwards, while there is a corresponding increase of erroneous identifications. The $10 h^{-1}$ Mpc filtered shows an increase of correct identifications at a very close range. This is a consequence of the rapidly rising ability to outline the corresponding larger underdensities as we expand towards a larger volume starting from the small nearby cosmic volume.

Filter scale and void finding

The DTFE reconstruction has been studied in somewhat more detail in Fig. 26. It plots the percentage of correctly and incorrectly identified void segments at a range of filter scales, running from $R_f = 1.0 h^{-1}$ Mpc, $R_f = 3.0 h^{-1}$ Mpc, $R_f = 6.0 h^{-1}$ Mpc to $R_f = 10.0 h^{-1}$ Mpc.

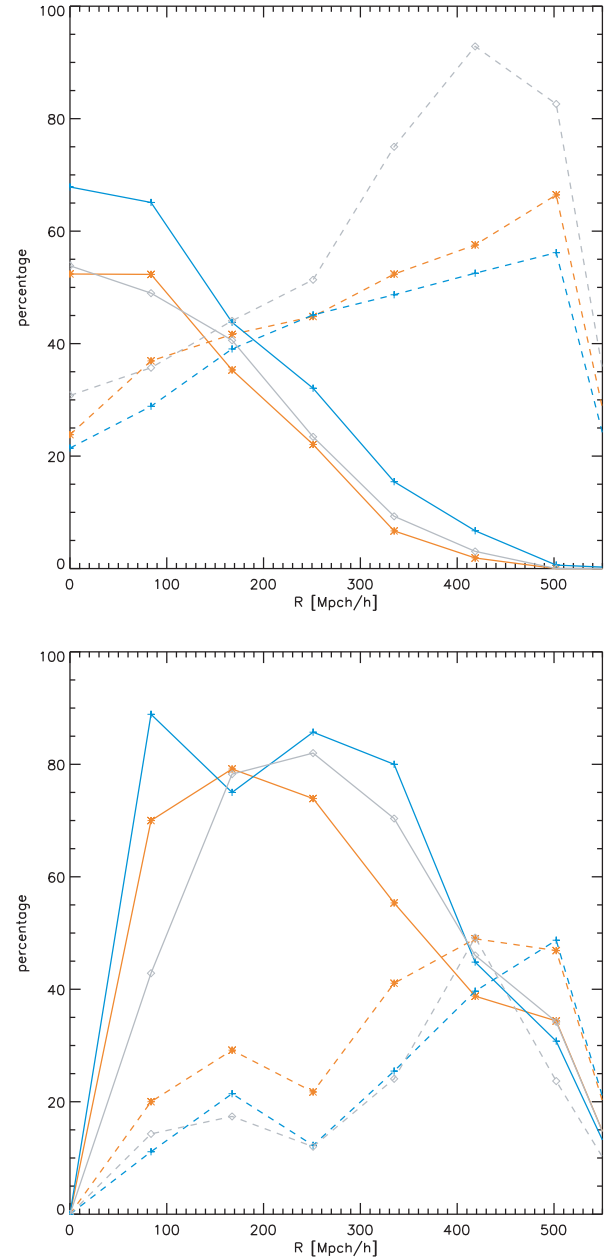


Figure 25. Correct and erroneous void identifications. Solid lines: (radially averaged) percentage of correct void patch identifications as a function of distance R . Dashed lines: (radially averaged) percentage of erroneous void patch identifications. Blue: DTFE; orange: NNFE; grey: Kriging. Top: density field reconstructions smoothed at $R_f = 3.0 h^{-1}$ Mpc. Bottom: density field reconstructions at $R_f = 10.0 h^{-1}$ Mpc.

For the larger filter scales, we find that the topology is reliably reconstructed within a range of $R \approx 100\text{--}300 h^{-1}$ Mpc, while for the smaller filter scales this seems to be at a much closer range from $R \approx 0$ to $200 h^{-1}$ Mpc. It is also interesting to see that on small scales, over the whole radial range, the number of correct identifications remains rather low and hardly exceeds 60 per cent. On filter scales of 6 and $10 h^{-1}$ Mpc the performance is certainly better and consistently acceptable over a $200 h^{-1}$ Mpc range. Only beyond $R \sim 300\text{--}400 h^{-1}$ Mpc the number of correct identifications drops below the 50 per cent mark.

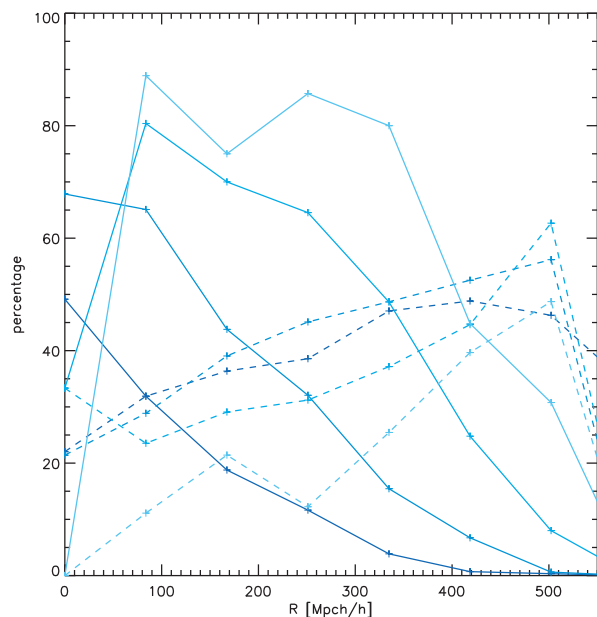


Figure 26. Correct and erroneous void identifications by DTFE. Solid lines: (radially averaged) percentage of correct void patch identifications as a function of distance R . Dashed lines: (radially averaged) percentage of erroneous void patch identifications. Various line textures indicate a range of filter scales: $R_f = 1.0 h^{-1}$ Mpc (dark blue), $3.0 h^{-1}$ Mpc, $6 h^{-1}$ Mpc and $10 h^{-1}$ Mpc (lightest blue).

7.6 Topology range

In general, we find that the number of correctly identified voids decreases at larger distances, corresponding with a related increase of topological errors with distance. Dependent on the filter scale, we may define a distance R_{\max} out to which DTFE/WVF manages to identify a reasonable amount of voids from a magnitude-limited survey. From Fig. 26 we find

- (i) $R_f = 1 h^{-1}$ Mpc: $R_{\max} \approx 100 h^{-1}$ Mpc
- (ii) $R_f = 3 h^{-1}$ Mpc: $R_{\max} \approx 200 h^{-1}$ Mpc
- (iii) $R_f = 6 h^{-1}$ Mpc: $R_{\max} \approx 300 h^{-1}$ Mpc
- (iv) $R_f = 10 h^{-1}$ Mpc: $R_{\max} \approx 400 h^{-1}$ Mpc.

Within this range, the fraction of erroneous void identifications remains well below unity: R_{\max} is roughly estimated from Fig. 26 from the scale R at which the solid curves (correct identification) and dashed curves (erroneous identification) cross.

Also, overall we find a slightly better performance by the DTFE reconstruction. Even though the higher order NNFE and Kriging techniques produce smoother void regions, this does not result in a higher number of correctly identified void patches.

8 SDSS-DR6 DENSITY RECONSTRUCTION

Following the extensive analysis discussed in the previous sections, we arrive at the application of the assessed technologies to the real world galaxy distribution in the sixth data release of SDSS. These resulting density maps form the starting point of the extensive statistical study – starting with a study of the one-point density distribution function and galaxy bias – presented in the subsequent papers discussing the density field and its cosmography.

The DTFE, NNFE and Kriging density field reconstructions based on the (magnitude-limited) galaxy distribution in the sixth data release of SDSS are shown in Fig. 28. The density field has

been reconstructed on a 512^3 grid, representing a spatial resolution of $1.17 h^{-1}$ Mpc.

For reference, in Fig. 27 we show the spatial distribution of the SDSS DR7 galaxies on which the reconstruction is based. They are superimposed as black dots on the DTFE density field.

Besides some minor differences, all three density maps show the same prominent features. The higher order NNFE and Kriging maps look smoother than the DTFE map. It is easy to recognize the almost one-to-one correspondence between the DTFE and NNFE map, with the DTFE map having the appearance of more noisy version. The Kriging map not only contains the same structures and features, but also has a more coherent appearance in which we can more easily recognize the global web-like morphology of the density field. Both the DTFE and NNFE maps have a more fragmented appearance.

9 SUMMARY AND DISCUSSION

This study is the first in a series in which we analyse the structure and topology of the Cosmic Web as traced by the SDSS.

In this study we investigate the ability of three reconstruction techniques to analyse and investigate web-like features and geometries in a discrete distribution of objects. The three methods are the DTFE, NNFE and a local ‘natural’ implementation of Kriging, the *Natural Lognormal Kriging* technique. DTFE and NNFE are based on the local geometry defined by the Voronoi and Delaunay tessellations of the galaxy distribution. The Kriging formalism is adapted and optimized for the approximate lognormal density distribution encountered in the mildly non-linear Cosmic Web and is based on the logarithm of the measured density values. Also, we have chosen to restrict the evaluations to a localized neighbourhood, the *tetrahedral natural neighbourhood* based on the Delaunay triangulation of the point sample.

The three reconstruction methods are analysed and compared using mock magnitude-limited and volume-limited SDSS redshift surveys, obtained on the basis of the Millennium simulation. We investigate error trends, biases and the topological structure of the resulting fields. The reconstructed density fields, mainly not only from the magnitude-limited mock survey samples but also from volume-limited ones, are compared with the density field of the total simulation galaxy sample. The differences between the various field reconstructions are investigated on the basis of an error analysis, mostly involving point-to-point comparisons. Environmental effects are addressed by evaluating the density fields on a range of Gaussian filter scales. With respect to the topology of the survey fields, we concentrate on the void population identified by the WVF (Platen et al. 2007). The void population in the full density field and in the magnitude-limited survey density fields is compared. The number of *false mergers* – in which original voids emerge as a part of a larger void in the survey field – and of *false splits* – in which an original void splits up in one or more voids in the survey field – forms the basis of the topological quality evaluation.

By investigating the quality of the resulting density field estimates, over a range of scales and in different environments, as well as the more global topological structure of the web-like network, we wish to identify and understand the qualities of these techniques for the different purposes and post-processing steps which are the subject of the following papers in this series. The following observations were made.

- (i) In most tests, DTFE, NNFE and Kriging have largely similar density and topology error behaviour.

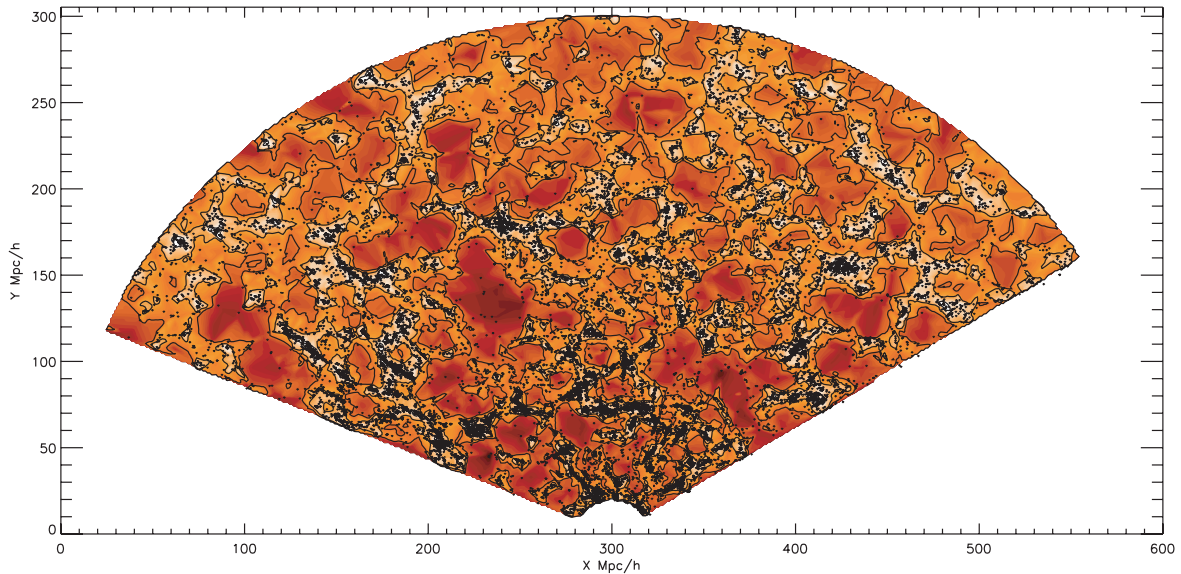


Figure 27. Galaxy distribution SDSS DR6. The magnitude-limited galaxy sample of the SDSS DR6 survey is superimposed on the DTFE density field contours. The density field section is infinitely thin, the galaxies lie within a slice of $10 h^{-1}$ Mpc width. The red coloured contour levels represent the underdense regions, at $\rho/\rho_u = [0.001, 0.002, 0.005, 0.01, 0.02, 0.03, 0.05, 0.07, 0.1, 0.2, 0.3, 0.6, 0.7, 0.8, 1]$. The heavy black solid line is the mean cosmic density level, $\delta = 0$. Within its contour are the overdense regions, marked by black contour lines at density $\rho/\rho_u = [1, 3, 10]$.

(ii) Cosmetically, higher order NNFE and Kriging methods produce more visually appealing reconstructions.

(iii) Quantitatively, DTFE performs (marginally) better. Part of this at first sight surprising finding is a consequence of the higher sensitivity of the higher order NNFE and Kriging interpolation to intrinsic errors in the galaxy sample. An additional factor is the smaller natural neighbourhood of DTFE and NNFE with respect to the three to four times larger neighbourhood of Natural Lognormal Kriging, which restricts density errors to smaller volumes.

(iv) With respect to the topological properties of the reconstructed density fields, it has become clear that a successful recovery of the void population on small scales is rather difficult. On these scales, the removal of only a couple of void galaxies leads to the spurious merging of observed voids.

(v) The void recovery rate improves significantly at filter scales $> 3 h^{-1}$ Mpc. The immediate repercussion is that a proper analysis of small-scale voids, and void galaxies, has to be necessarily restricted to the local Universe out to at most $100 h^{-1}$ Mpc. As the environmental influences on the galaxy formation process seem to be mostly determined on these scales (Park et al. 2007), our study within this project, subject of a forthcoming paper in this series, will restrict itself mainly to this volume.

A variety of technical improvements of our DTFE, NNFE and Kriging implementations may lead to a better performance. One immediate option is to invoke an image grid which has a more natural character for the galaxy survey context of our study than the cubic grid we have used in the evaluations described in this paper. Because of the flexibility of their definition, the Kriging formalism will be a promising context for further adaptations and improvements. Of immediate importance is the implementation of non-local techniques to optimize the matrix inversion calculations while retaining the influence of large-scale correlations and predictor-corrector methods to deal with oscillatory instabilities. While we have not yet investigated its performance, results of studies in other fields emphasize

the importance of evaluating the performance of RBF techniques as a possible alternative.

The DTFE, NNFE and Kriging density field reconstructions form the basis of a series of studies in which we analyse the cosmography, void population and spinal structure of the local Cosmic Web in the SDSS. Given the ease and efficiency of calculation, and its good quantitative behaviour, for most of these studies we use the DTFE formalism. However, the Natural Lognormal Kriging results look very promising and appear to produce a well-behaved coherent web-like density map of the SDSS survey. With the large advantage of controlling the error behaviour and properties of the reconstructed map, along with the large potential for extensions and optimizations of the method, the Natural Lognormal Kriging maps will play a dominant role in our study of the Cosmic Web.

The density field reconstructions within the SDSS DR6, and subsequently SDSS DR7, volumes will be subjected to a statistical study in the second paper of this series. We will focus in particular on the one-point probability function of the SDSS density field. With the help of corresponding magnitude-limited mock catalogues we will infer in how far the lognormal probability function, as well as related higher order versions, form a proper description of its statistical character. Also, the mock catalogues will allow us to assess the bias of the galaxy population in high-density areas and, in particular, the low-density void region. A cosmographic description of the reconstructed Local Universe, along with the identification of filamentary supercluster complexes, voids and supervoid complexes, is the subject of the third paper in this series. The study of the void population in the SDSS density field, concerning a detailed assessment of the void size and shape properties, forms an important rationale behind the development of the density field reconstructions described in this study. This has become an interesting area of research, in particular as recent studies have emphasized the potential of extracting cosmological information from cosmic voids, in particular that concerning the dark energy equation of state (Lee & Park 2009; Lavaux & Wandelt 2010; Biswas, Alizadeh & Wandelt 2010). In recent years, there

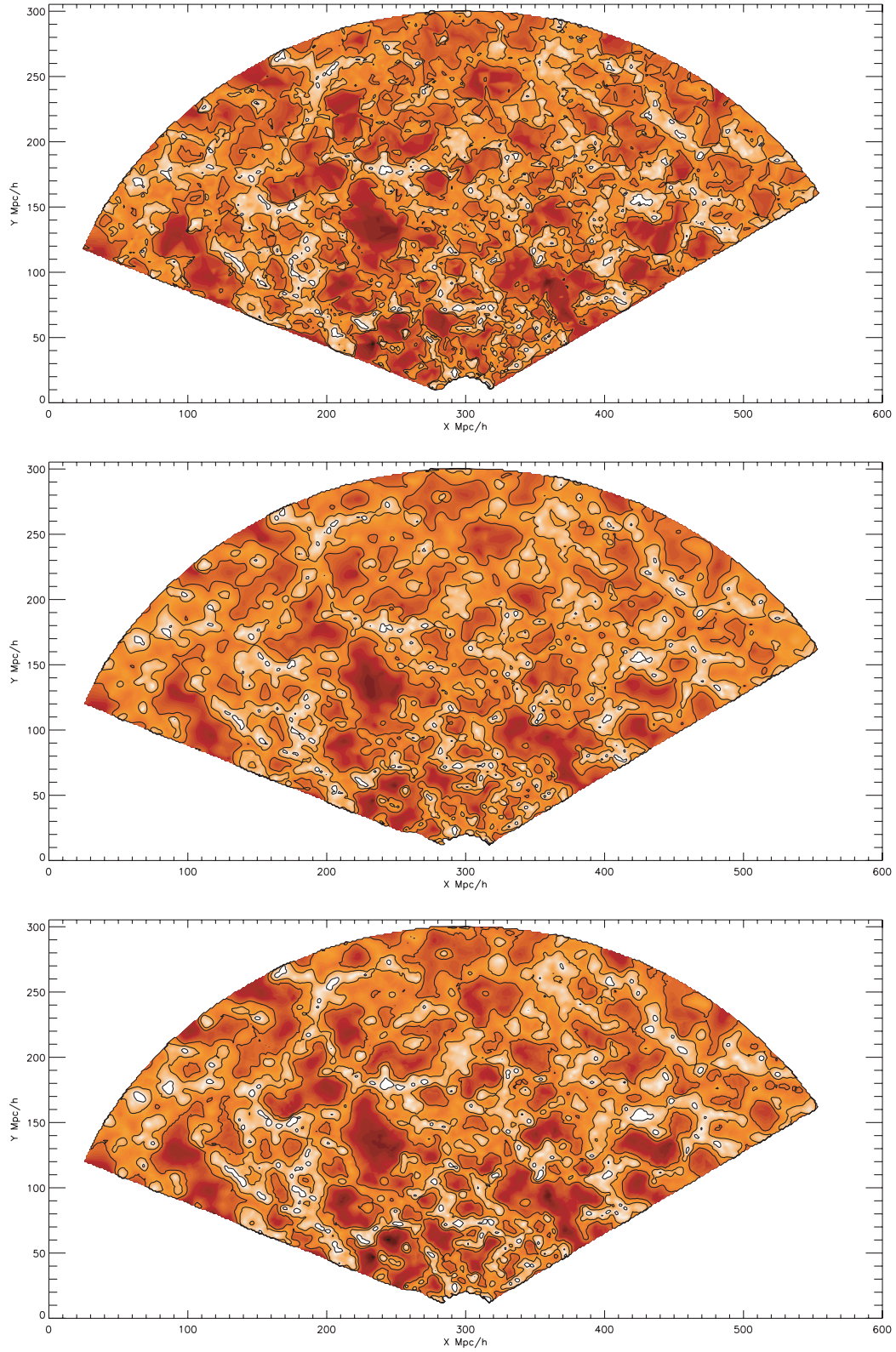


Figure 28. SDSS DR6 density field reconstructions. Top: DTFE; centre: NNFE; bottom: Kriging. Shown are thin – zero width – slice sections through the reconstructed density field. The red coloured contour levels represent the underdense regions, at $\rho/\rho_u = [0.001, 0.002, 0.005, 0.01, 0.02, 0.03, 0.05, 0.07, 0.1, 0.2, 0.3, 0.6, 0.7, 0.8, 1]$. The heavy black solid line is the mean cosmic density level, $\delta = 0$. Within its contour are the overdense regions, marked by black contour lines at density $\rho/\rho_u = [1, 3, 10]$.

has also been a strong interest in large-scale environmental influences on galaxy properties and on the galaxy formation process. The tools described in the present study will allow an assessment on the basis of a properly defined density field on quasi-linear scales.

ACKNOWLEDGMENTS

EP wishes to thank Alex Szalay for his hospitality and financial support at JHU. We are also thankful for the help from Gerard Lemson and Sebastien Heinis with the Millennium Data base queries. We would like to thank Istvan Szapudi and Mark Neyrinck for their helpful discussion. We are grateful to the referee, Michael Vogeley, for the many incisive and helpful remarks and recommendations.

The Millennium Simulation databases used in this paper and the web application providing online access to them were constructed as part of the activities of the German Astrophysical Virtual Observatory (GAVO).

Funding for the SDSS and SDSS-II has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, the US Department of Energy, the National Aeronautics and Space Administration, the Japanese Monbukagakusho, the Max Planck Society and the Higher Education Funding Council for England. The SDSS Web Site is <http://www.sdss.org/>.

The SDSS is managed by the Astrophysical Research Consortium for the Participating Institutions. The Participating Institutions are the American Museum of Natural History, Astrophysical Institute Potsdam, University of Basel, University of Cambridge, Case Western Reserve University, University of Chicago, Drexel University, Fermilab, the Institute for Advanced Study, the Japan Participation Group, Johns Hopkins University, the Joint Institute for Nuclear Astrophysics, the Kavli Institute for Particle Astrophysics and Cosmology, the Korean Scientist Group, the Chinese Academy of Sciences (LAMOST), Los Alamos National Laboratory, the Max-Planck-Institute for Astronomy (MPIA), the Max-Planck-Institute for Astrophysics (MPA), New Mexico State University, Ohio State University, University of Pittsburgh, University of Portsmouth, Princeton University, the United States Naval Observatory and the University of Washington.

REFERENCES

- Abgrall R., 1994, *J. Comput. Phys.*, 114, 45
 Adelman-McCarthy J. et al., 2008, *ApJS*, 175, 297
 Amidror I., 2002, *J. Electron. Imaging*, 11, 157
 Aragón-Calvo M. A., Jones B. J. T., van de Weygaert R., van der Hulst J. M., 2007, *A&A*, 474, 315
 Aragón-Calvo M. A., van de Weygaert R., Platen E., Szalay A. S., 2010, *ApJ*, 723, 364
 Arbabi-Bidgoli S., Müller V., 2002, *MNRAS*, 332, 205
 Ascasibar Y., Binney J., 2005, *MNRAS*, 356, 872
 Aurenhammer F., 1991, *ACM Computing Surveys*, 23, 345
 Bahcall N., Soneira R., 1982, *ApJ*, 262
 Bernardeau F., van de Weygaert R., 1996, *MNRAS*, 279, 693
 Bertone S., De Lucia G., Thomas P. A., 2007, *MNRAS*, 379, 1143
 Bertschinger E., 1987, *ApJ*, 323
 Biswas R., Alizadeh E., Wandelt B. D., 2010, *Phys. Rev. D*, 82, 023002
 Blaizot J., Wadadekar Y., Guiderdoni B., Colombi S., Bertin E., Bouchet F., Devriendt J., Hatton S., 2005, *MNRAS*, 360, 159
 Blanton M. R., Lin H., Lupton R. H., Maley F. M., Young N., Zehavi I., Loveday J., 2003, *AJ*, 125, 2276
 Bond J., Kofman L., Pogossyan D., 1996, *Nat*, 380, 603
 Bower R. G., Benson A. J., Malbon R., Helly J. C., Frenk C. S., Baugh C. M., Cole S., Lacey C. G., 2006, *MNRAS*, 370, 645
 Braun J., Sambridge M., 1995, *Nat*, 376, 655
 Brown G., 1965, *New Zealand Forestry Service Res. Notes*, 38, 1
 Browne M., 2007, *Pattern Recognition*, 40, 134
 Colberg J. M., Sheth R. K., Diaferio A., Gao L., Yoshida N., 2005, *MNRAS*, 360, 216
 Colberg J. et al., 2008, *MNRAS*, 387, 933
 Coles P., Jones B., 1991, *MNRAS*, 248, 1
 Colless M. et al., 2003, preprint (astro-ph/0306581)
 Computational Geometry Algorithms Library, CGAL, <http://www.cgal.org>
 Cressie N., 1988, *Math. Geology*, 20, 405
 Cressie N., 1990, *Math. Geology*, 22, 239
 Cressie N., 1993, *Statistics for Spatial Data*, revised edition, Vol. 605. Wiley, New York
 De Lucia G., Blaizot J., 2007, *MNRAS*, 375, 2
 Delaunay B., 1934, *Bull. Acad. Sci. USSR (VII) Classe Sci. Mat.*, p. 793
 Dressler A., 1980, *ApJ*, 236, 351
 Ebeling H., Wiedenmann G., 1993, *Phys. Rev. E*, 47, 704
 Efsthathiou G., Moody S., 2001, *MNRAS*, 325, 1603
 Einasto J. et al., 2007, *A&A*, 462, 811
 El-Ad H., Piran T., 1997, *ApJ*, 491, 421
 Eldering B., Vegter G., Weygaert R., Kruithof N., 2006, Master thesis, University of Groningen
 Erdogdu P. et al., 2004, *MNRAS*, 352, 939
 Erdogdu P. et al., 2006, *MNRAS*, 373, 45
 Franke R., 1982, *Math. Computation* 38, 181
 Genovese C. R., Perone-Pacifico M., Verdini I., Wasserman L., 2010, preprint (arXiv:1003.5536)
 Giovanelli R., Haynes M. P., Chincarini G. L., 1986, *ApJ*, 300, 77
 Gott J. R., Dickinson M., Melott A. L., 1986, *ApJ*, 306, 341
 Gott J. R. et al., 1989, *ApJ*, 340, 625
 Gott J. R. et al., 2008, *ApJ*, 675, 16
 Hockney R. W., Eastwood J. W., 1981, *Computer Simulation Using Particles*. McGraw-Hill, New York
 Hoekstra H., Jain B., 2008, *Annu. Rev. Nuclear Part. Sci.*, 58, 99
 Hoffman Y., Ribak E., 1991, *ApJ*, 380, 5
 Hoyle F., Vogeley M. S., Gott J. R. I., 2002, *ApJ*, 570, 44
 Huchra J. et al., 2005, in Fairall A. P., Woudt P. A., eds, *ASP Conf. Ser. Vol. 329. Nearby Large-Scale Structures and the Zone of Avoidance. The 2MASS Redshift Survey and Low Galactic Latitude Large-Scale Structure*. Astron. Soc. Pac., San Francisco, p. 135
 Katkovnik V., Shmulevich I., 2002, *Pattern Recognition Lett.*, 23, 1641
 Kauffmann G., Fairall A. P., 1991, *MNRAS*, 248, 313
 Kayo I., Taruya A., Suto Y., 2001, *ApJ*, 561, 22
 Kirshner R. P., Oemler A., Jr, Schechter P. L., Shectman S. A., 1981, *ApJ*, 248, L57
 Kitaura F. S., Enßlin T. A., 2008, *MNRAS*, 389, 497
 Kitaura F. S., Jasche J., Li C., Enßlin T. A., Metcalf R. B., Wandelt B. D., Lemson G., White S. D. M., 2009, *MNRAS*, 400, 183
 Kitaura F., Jasche J., Metcalf R. B., 2010, *MNRAS*, 403, 589
 Kofman L., Bertschinger E., Gelb J., Nusser A., Dekel A., 1994, *ApJ*, 420
 Krige D., 1951, *J. Chem Metall. Min. Soc. S. Africa*, 52, 119
 Lavaux G., Wandelt B. D., 2010, *MNRAS*, 403, 1392
 Lee J., Park D., 2009, *ApJ*, 696, L10
 Lloyd S., 1982, *Inf. Theor., IEEE Trans.*, 28, 129
 Lombardi M., 2002, *A&A*, 395, 733
 Martínez V. J., Starck J.-L., Saar E., Donoho D. L., Reynolds S. C., de la Cruz P., Paredes S., 2005, *ApJ*, 634, 744
 Massey R. et al., 2007, *Nat*, 445, 286
 Matheron G., 1963, *Economic Geology*, 58, 1246
 Mecke K. R., Buchert T., Wagner H., 1994, *A&A*, 288, 697
 Mellier Y., 1999, *ARA&A*, 37, 127
 Miller E. G., 2003, *A New Class of Entropy Estimators for Multi-dimensional Densities*. Proc. IEEE Int. Conf. on Acoustics, Speech & Signal Proc. (ICASSP), 3, p. 297
 Monaghan J. J., 1992, *ARA&A*, 30, 543
 Neyrinck M. C., 2008, *MNRAS*, 386, 2101

- Okabe A., Boots B., Sugihara K., Chiu S. N., 2000, *Spatial Tessellations, Concepts and Applications of Voronoi Diagrams*. John Wiley, Chichester
- Ord J., 1978, *Math. Scientist*, 3, 23
- Park C., Gott J. R., Melott A. L., Karachentsev I. D., 1992, *ApJ*, 387, 1
- Park C., Choi Y., Vogeley M., Gott J., III, Blanton M., 2007, *ApJ*, 658, 898
- Patiri S. G., Betancort-Rijo J. E., Prada F., Klypin A., Gottlöber S., 2006, *MNRAS*, 369, 335
- Peebles P., 1980, *The Large Scale Structure of the Universe*. Princeton Univ. Press, Princeton
- Percival W. et al., 2001, *MNRAS*, 327, 1297
- Platen E., van de Weygaert R., Jones B., 2007, *MNRAS*, 380, 551
- Plionis M., Basilakos S., 2002, *MNRAS*, 330, 399
- Press W. H. a., 2007, *Numerical Recipes: The Art of Scientific Computing*, Third Edition. Cambridge Univ. Press, Cambridge
- Rybicki G., Press W., 1992, *ApJ*, 398, 169
- Sahni V., Sathyaprakash B. S., Shandarin S. F., 1998, *ApJ*, 495, L5
- Sain S. R., 2002, *Comp. Stat. Data Anal.*, 39, 165
- Schaap W. E., 2007, PhD thesis, Univ. Groningen
- Schaap W. E., van de Weygaert R., 2000, *A&A*, 363, L29
- Schmalzing J., Buchert T., Melott A. L., Sahni V., Sathyaprakash B. S., Shandarin S. F., 1999, *ApJ*, 526, 568
- Shandarin S., Sheth J., Sahni V., 2004, *MNRAS*, 353, 162
- Shandarin S., Feldman H. A., Heitmann K., Habib S., 2006, *MNRAS*, 376, 1629
- Sharma S., Steinmetz M., 2006, *MNRAS*, 373, 1293
- Sheth R., 1995, *MNRAS*, 277, 933
- Sibson R., 1981, Barnet V., ed. *Interpreting Multi-variate Data*. Wiley, Chichester
- Silverman B. W., 1986, *Density Estimation for Statistics and Data Analysis*, Monograph Statist. and Appl. Prob. Chapman & Hall, London
- Smoot G. et al., 1992, *ApJ*, 396, L1
- Sousbie T., 2011, *MNRAS*, 414, 350
- Spergel D. et al., 2003, *ApJS*, 148, 175
- Springel V. et al., 2005, *Nat*, 435, 629
- Stoughton C. et al., 2002, *AJ*, 123, 485
- Strauss M. et al., 2002, *AJ*, 124, 1810
- Sukumar N., 1998, PhD thesis, Northwestern Univ.
- Sukumar N., Moran B., Belytschko T., 1998, *Int. J. Numer. Meth. Eng.*, 43, 839
- Tegmark M. et al., 2004, *ApJ*, 606, 702
- van de Weygaert R., Bertschinger E., 1996, *MNRAS*, 281, 84
- van de Weygaert R., Schaap W. E., 2008, in Martínez V., Saar E., Martínez-González E., Pons-Bordería M., eds. *Lecture Notes in Physics No. 665*, Data Analysis in Cosmology. Springer-Verlag, Berlin, p. 289
- Watson D. F., 1992, *Contouring: A Guide to the Analysis and Display of Spatial Data*. Pergamon Press, New York
- Way M. J., Gazis P. R., Scargle J. D., 2011, *ApJ*, 727, 48
- Wiener N., 1949, *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*. The MIT Press, Cambridge, MA
- York D. et al., 2000, *AJ*, 120, 1579
- Zaroubi S., 2002, *MNRAS*, 331, 901
- Zaroubi S., Hoffman Y., Fisher K. B., Lahav O., 1995, *ApJ*, 449, 446
- Zehavi E. et al., 2010, *ApJ*, 236, 351
- Zel'Dovich Y. B., 1970, *A&A*, 5, 84
- Zhang Y., Springel V., Yang X., 2009, *ApJ*, 722, 812

APPENDIX A: SDSS COORDINATE SYSTEM

For the analysis of our data sample, we transform the equatorial coordinates (α, δ, z) of the DR6 NGP and DR7 galaxy sample to a grid-based coordinate system (X, Y, Z) . The observer is located at $(X, Y, Z) = (300, 0, 300) h^{-1} \text{ Mpc}$, while the centre of the northern strip is rotated to lie parallel to the Y -axis, starting at $(X, Z) = (300, 300) h^{-1} \text{ Mpc}$. The corresponding transformation for an object with

a comoving distance $R(z) = cz/100 h^{-1} \text{ Mpc}$ is defined by

$$\begin{aligned} X &= R(z) \cos(\delta) \cos(\alpha - 90) \\ Y &= R(z) \cos(\delta) \sin(\alpha - 90) \\ Z &= R(z) \sin(\delta). \end{aligned} \quad (\text{A1})$$

APPENDIX B: DTFE: DELAUNAY TESSELLATION FIELD ESTIMATOR

An extensive outline of the full DTFE procedure can be found in van de Weygaert & Schaap (2008). For the specific application to the SDSS density field reconstruction, we follow the following steps of the DTFE procedure:

(1) Point sample

The (mock) galaxy samples are supposed to represent an unbiased sample of the underlying density field. It is therefore considered to be a general Poisson process of the underlying density field.

(2) Boundary conditions

We assume *vacuum boundary conditions*: outside the galaxy sample volume we take the minimal assumption of having no points.

(3) Delaunay tessellation

Construct the Delaunay tessellation from the point sample using the Computational Geometry Algorithms Library (CGAL) library.

(4) Field values point sample

The density values at the sampled points are determined from the corresponding Voronoi tessellations. The estimate of the density at each sample point is the normalized inverse of the volume of its *contiguous* Voronoi cell \mathcal{W}_i of each point i . The *contiguous Voronoi cell* of a point i is the union of all Delaunay tetrahedra of which point i forms one of the four vertices (see Fig. 5 for an illustration). We recognize two applicable situations:

(i) Uniform sampling process

The point sample is an unbiased sample of the underlying density field. Typical example is that of N -body simulation particles. For D -dimensional space the density estimate is

$$\hat{f}(\mathbf{x}_i) = (1 + D) \frac{m_i}{V(\mathcal{W}_i)} \quad (\text{B1})$$

with m_i the ‘mass’ of sample point i . This situation concerns the ‘full’ mock galaxy samples and the volume-limited galaxy samples.

(ii) Systematic non-uniform sampling process

Sampling density according to the specified selection process quantified by an a priori known selection function $\psi(\mathbf{x})$, varying as a function of sky position (α, δ) as well as depth/redshift. For D -dimensional space the density estimate is

$$\hat{f}(\mathbf{x}_i) = (1 + D) \frac{m_i}{\psi(\mathbf{x}_i) V(\mathcal{W}_i)}. \quad (\text{B2})$$

This situation is relevant for the magnitude- or flux-limited SDSS and mock galaxy samples.

(5) Field gradient

Calculation of the field gradient estimate $\widehat{\nabla} f|_m$ in each D -dimensional Delaunay simplex m ($D = 3$: tetrahedron; $D = 2$: triangle) by solving the set of linear equations for the field values at the positions of the $(D + 1)$ tetrahedron vertices,

$$\widehat{\nabla} f|_m \leftarrow \begin{Bmatrix} f_0 & f_1 & f_2 & f_3 \\ \mathbf{r}_0 & \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{r}_3. \end{Bmatrix} \quad (\text{B3})$$

(6) *Interpolation*

The final basic step of the DTFE procedure is the field interpolation. The processing and post-processing steps involve numerous interpolation calculations, for each of the involved locations $\hat{\mathbf{r}}$. Given a location $\hat{\mathbf{r}}$, the Delaunay tetrahedron m in which it is embedded is determined. On the basis of the field gradient $\widehat{\nabla}f|_m$ the field value is computed by (linear) interpolation,

$$\hat{f}(\mathbf{r}) = \hat{f}(\mathbf{r}_i) + \widehat{\nabla}f|_m \cdot (\mathbf{r} - \mathbf{r}_i). \quad (\text{B4})$$

(7) *Processing*

We make a distinction between straightforward processing steps concerning the production of images and simple smoothing filtering operations on the one hand, and more complex post-processing on the other hand. Basic to the processing steps is the determination of field values following the interpolation procedure(s) outlined above. Straightforward ‘first line’ field operations are ‘*Image reconstruction*’ and, subsequently, ‘*Smoothing/Filtering*’.

(i) *Image reconstruction*

For a set of *image points*, usually grid points, determine the *image value*: formally the average field value within the corresponding grid cell. For that purpose in this study we use the *Monte Carlo approach*: approximate the integral by taking the average over a number of (interpolated) field values probed at randomly distributed locations within the grid cell around an *image point*. The final estimate is obtained by averaging over the interpolated field values within a grid cell.

For image reconstruction we need to assure ourselves that we obtain a sensible density estimate within a voxel element. In a spatially irregular sample, the Voronoi cell defines the natural Nyquist interval. To avoid aliasing, the number of interpolation points therefore needs to oversample the voxels of the image grid. One may also take the alternative and exact option of piecewise integrating the density of the Delaunay tetrahedra that are (partially) overlapping with the image voxel. For DTFE this can be accomplished exactly and relatively fast, although the computational and geometric aspects are far from trivial.

(ii) *Smoothing and filtering*

Linear filtering of the field \hat{f} : convolution of the field \hat{f} with a filter function $W_f(\mathbf{r}, \mathbf{y})$, usually user-specified,

$$f_s(\mathbf{r}) = \int \hat{f}(\mathbf{r}') W_f(\mathbf{r}', \mathbf{y}) d\mathbf{r}' \quad (\text{B5})$$

(8) *Post-processing*

The real potential of DTFE fields may be found in sophisticated applications, tuned towards uncovering characteristics of the reconstructed fields. An important aspect of this involves the analysis of structures in the density field. This can be finding voids, identifying cosmic structures or advanced statistical analysis of the density field.

This paper has been typeset from a $\text{\TeX}/\text{\LaTeX}$ file prepared by the author.